

THE JOURNAL of the Acoustical Society of America

Vol. 101, No. 6

June 1997

| | | |
|--|---|------|
| ACOUSTICAL NEWS—USA | | 3221 |
| USA Meetings Calendar | | 3224 |
| ACOUSTICAL NEWS—INTERNATIONAL | | 3228 |
| International Meetings Calendar | | 3228 |
| REVIEWS OF ACOUSTICAL PATENTS | | 3230 |
| SELECTED RESEARCH ARTICLE [10] | | |
| Speech perception as pattern recognition | Terrance M. Nearey | 3241 |
| GENERAL LINEAR ACOUSTICS [20] | | |
| Goupillaud inverse problem with arbitrary input | Wang Ning, Lin JunXuan, Ueha Sadayuki | 3255 |
| On the scattering of antisymmetric edge modes | M. de Billy | 3261 |
| A generalized internal source density method for the forward and backward projection of harmonic pressure fields from complex bodies | Peter R. Stepanishen | 3270 |
| Ultrasonic anisotropic phase velocity determination with the Radon transformation | A. Kebaili, D. R. Schmitt | 3278 |
| An efficient finite element scheme for solving the three-dimensional poroelasticity problem in acoustics | Raymond Panneton, Noureddine Atalla | 3287 |
| Acoustic dispersion and attenuation in many spherical scatterer systems and the Kramers–Kronig relations | Zhen Ye | 3299 |
| Reduction of structure-borne noise using an air-voided elastomer | Sung H. Ko | 3306 |
| A combined integro-modal approach for predicting acoustic properties of irregular-shaped cavities | J. Missaoui, L. Cheng | 3313 |
| Effects of an elastic membrane on tube waves in permeable formations | Hsui-Lin Liu, David Linton Johnson | 3322 |
| A boundary integral method for acoustic radiation and scattering | W. S. Hwang | 3330 |
| Hypersingular boundary integral equations for exterior acoustic problems | W. S. Hwang | 3336 |
| AEROACOUSTICS, ATMOSPHERIC SOUND [28] | | |
| Sound propagation from a dipole source near an impedance plane | Kai Ming Li, Shahram Taherzadeh, Keith Attenborough | 3343 |
| Oscillation modes of supersonic multijets | Y. Umeda, R. Ishii | 3353 |
| UNDERWATER SOUND [30] | | |
| Frequency dependence of broadband propagation in coastal regions | M. Badiey, J. Simmen, S. Forsythe | 3361 |

(Continued)

CONTENTS—Continued from preceding page

| | | |
|---|--|------|
| All-frequency normal-mode solution of the three-dimensional acoustic scattering from a vertical cylinder in a plane-horizontal waveguide | Gerassimos A. Athanassoulis, Konstadinos A. Belibassakis | 3371 |
| Observations on attenuation and shear-wave velocity in fine-grained, marine sediments | Frederick A. Bowles | 3385 |
| Direct laboratory measurement of forward scattering by individual fish | Li Ding | 3398 |
| ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND [35] | | |
| Anomalous attenuation effect on reflectivity of an ultrasonic wave from a thin layer between dissimilar materials | A. I. Lavrentyev, S. I. Rokhlin | 3405 |
| Bounds and approximations for elastic wave speeds in cubic crystals | Q. H. Zuo, K. D. Hjelmstad | 3415 |
| Visualization of acoustic particle interaction and agglomeration: Theory evaluation | Thomas L. Hoffmann, Gary H. Koopmann | 3421 |
| STRUCTURAL ACOUSTICS AND VIBRATION [40] | | |
| Vibration and wave localization in a nearly periodic beaded string | Gisli Óttarsson, Christophe Pierre | 3430 |
| Diffuse transmission of structure-borne sound at periodic junctions of semi-infinite plates | I. Bosmans, G. Vermeir | 3443 |
| Frequency response of a two-dimensional trusslike periodic panel | Michael El-Raheb | 3457 |
| Multiple-scattering theory for mean responses in a plate with sprung masses | Richard L. Weaver | 3466 |
| Estimating acoustic radiation from a Bernoulli–Euler beam using shaped polyvinylidene fluoride film | Brian L. Scott, Scott D. Sommerfeldt | 3475 |
| NOISE: ITS EFFECTS AND CONTROL [50] | | |
| Comparison between subjective and objective measures of active hearing protector and communication headset attenuation | Jan Zera, Anthony J. Brammer, George J. Pan | 3486 |
| Generation of zones of quiet using a virtual microphone arrangement | J. Garcia-Bonito, S. J. Elliott, C. C. Boucher | 3498 |
| ARCHITECTURAL ACOUSTICS [55] | | |
| Characterization of a diffuse field in a reverberant room | H. Nélisse, J. Nicolas | 3517 |
| ACOUSTIC SIGNAL PROCESSING [60] | | |
| Matched-field localization for multiple sources in an uncertain environment, with application to Arctic ambient noise | Michael V. Greening, Pierre Zakarauskas, Stan E. Dosso | 3525 |
| Source localization in noisy and uncertain ocean environments | Laurie T. Fialkowski, Michael D. Collins, John S. Perkins, W. A. Kuperman | 3539 |
| PHYSIOLOGICAL ACOUSTICS [64] | | |
| Quantitative measures of hair cell loss in CBA and C57BL/6 mice throughout their life spans | Vlasta P. Spongr, Dorothy G. Flood, Robert D. Frisina, Richard J. Salvi | 3546 |
| Fine structure of the $2f_1-f_2$ acoustic distortion product: Effects of primary level and frequency ratios | Ning-ji He, Richard A. Schmiedt | 3554 |
| Are normal hearing thresholds a sufficient condition for click-evoked otoacoustic emissions? | Sarosh Kapadia, Mark E. Lutman | 3566 |
| Changes in evoked otoacoustic emissions in the guinea pig after pure-tone acoustic overstimulation | Hiromi Ueda, Hayato Tsuge, Taku Hattori | 3577 |
| The mechanical waveform of the basilar membrane. I. Frequency modulations (“glides”) in impulse responses and cross-correlation functions | Egbert de Boer, Alfred L. Nuttall | 3583 |

CONTENTS—Continued from preceding page

| | | |
|---|--|------|
| Fluid–structure interaction of the stereocilia bundle in relation to mechanotransduction | D. E. Zetes, C. R. Steele | 3593 |
| Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers | Roger L. Miller, John R. Schilling, Kevin R. Franck, Eric D. Young | 3602 |
| PSYCHOLOGICAL ACOUSTICS [66] | | |
| The modulated–unmodulated difference: Effects of signal frequency and masker modulation depth | Sid P. Bacon, Jungmee Lee | 3617 |
| Sensitivity to changes in overall level and spectral shape: An evaluation of a channel model | Jennifer J. Lentz, Virginia M. Richards | 3625 |
| Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system | Robert P. Carlyon, A. Jaysurya Datta | 3636 |
| Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components | Robert P. Carlyon, A. Jaysurya Datta | 3648 |
| Auditory filters measured at neighboring center frequencies | Marc A. Fagelson, Craig A. Champlin | 3658 |
| A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing | Andrew J. Oxenham, Christopher J. Plack | 3666 |
| Short-term temporal integration: Evidence for the influence of peripheral compression | Andrew J. Oxenham, Brian C. J. Moore, Deborah A. Vickers | 3676 |
| Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration | Jungmee Lee, Sid P. Bacon | 3688 |
| Detection of silent intervals between noises activating different perceptual channels: Some properties of “central” auditory gap detection | D. P. Phillips, T. L. Taylor, S. E. Hall, M. M. Carr, J. E. Mossop | 3694 |
| Psychometric functions and temporal integration in electric hearing | Gail S. Donaldson, Neal F. Viemeister, David A. Nelson | 3706 |
| SPEECH PRODUCTION [70] | | |
| Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa | Roger W. Chan, Ingo R. Titze, Michael R. Titze | 3722 |
| Articulatory strengthening at edges of prosodic domains | Cécile Fougeron, Patricia A. Keating | 3728 |
| Coarticulatory stability in American English /r/ | Suzanne Boyce, Carol Y. Espy-Wilson | 3741 |
| The characteristics of voicing in syllable-initial fricatives in American English | Karen Pirello, Sheila E. Blumstein, Kathleen Kurowski | 3754 |
| SPEECH PERCEPTION [71] | | |
| Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants | Margaret W. Skinner, Laura K. Holden, Timothy A. Holden, Marilyn E. Demorest, Marios S. Fourakis | 3766 |
| BIOACOUSTICS [80] | | |
| Propagation of damselfish (<i>Pomacentridae</i>) courtship sounds | David A. Mann, Phillip S. Lobel | 3783 |
| LETTERS TO THE EDITOR | | |
| Background contributions and coupling coefficients for backscattering by thick shells [20] | Steven G. Kargl, Philip L. Marston | 3792 |
| Exact solution for one-dimensional acoustic fields in ducts with a quadratic mean temperature profile [20] | B. Manoj Kumar, R. I. Sujith | 3798 |
| Comments on “Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I and Part II” [J. Acoust. Soc. Am. 99, 1339–1348, 1349–1359 (1996)] [20] | Jamal Assaad, Jean-Michel Rouvaen | 3800 |

CONTENTS—Continued from preceding page

| | | |
|--|---|------|
| Response to “Comments on ‘Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I and Part II’” [J. Acoust. Soc. Am. 101, 3800–3803 (1997)] [20] | Ping Wu, Rymantas Kazys, Tadeusz Stepinski | 3804 |
| Effect of monaural and binaural altered auditory feedback on stuttering frequency [70] | Andrew Stuart, Joseph Kalinowski, Michael P. Rastatter | 3806 |
| ERRATA | | |
| Erratum: “Vibrational impact of high-speed trains. I. Effect of track dynamics” [J. Acoust. Soc. Am. 100, 3121–3134 (1996)] | Victor V. Krylov | 3810 |
| TECHNICAL NOTES AND RESEARCH BRIEFS | | 3811 |
| INDEX TO VOLUME 101 | | 3812 |
| SUBJECT INDEX TO VOLUME 101 | | 3817 |
| AUTHOR INDEX TO VOLUME 101 | | 3849 |

NOTES CONCERNING ARTICLE ABSTRACTS

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

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

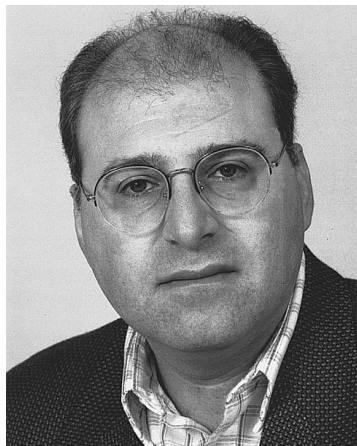
ACOUSTICAL NEWS—USA

Elaine Moran

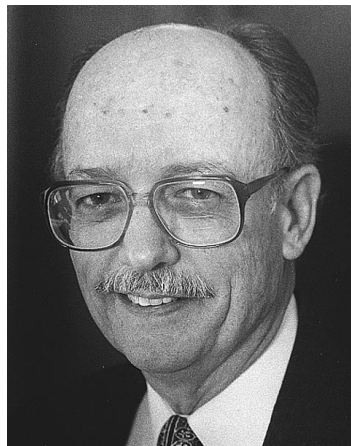
Acoustical Society of America, 500 Sunnyside Boulevard, Woodbury, New York 11797

Editor's Note: Deadline dates for news items and notices are 2 months prior to publication.

New Fellows of the Acoustical Society of America



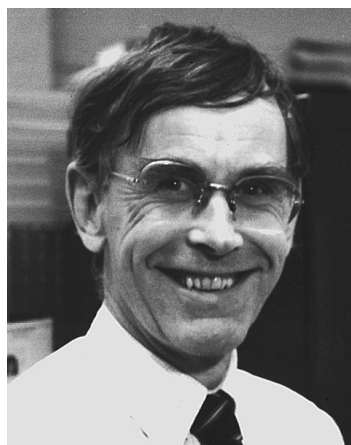
Leslie R. Bernstein—For contributions to binaural hearing.



Robert C. Coffeen—For contributions to electroacoustics in buildings.

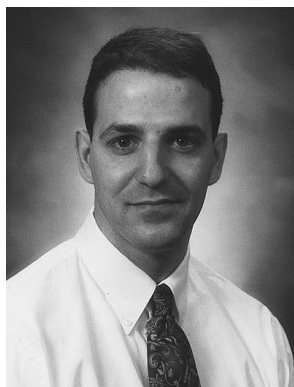


Jens-Peter Koester—For contributions to speaker verification and phonetic science.



Anders Lofqvist—For contributions to understanding the control mechanisms of speech production.

F. V. Hunt Postdoctoral Research Fellowship awarded to James J. Finneran



The 1997–1998 F. V. Hunt Postdoctoral Research Fellowship in Acoustics was awarded to James J. Finneran. Under the Fellowship Dr. Finneran will undertake a research program at the Naval Command, Control, and Ocean Surveillance Center in San Diego, CA. The subject of his research program is on the physiology of dolphin hearing.

Dr. Finneran holds B.S., M.S., and Ph.D. degrees from the Department of Mechanical Engineering, The Ohio State University. His Ph.D. thesis was on "Experimental and theoretical investigation of the low-frequency response of the goldfish peripheral auditory system."

The Hunt Fellowship is granted each year to an ASA member who has

recently received his or her doctorate or will be receiving the degree in the year in which the fellowship is to be granted. The recipient of the fellowship is that individual who, through personal qualifications and a proposed research topic, is judged to exhibit the highest potential for benefiting any aspect of the science of sound and promoting its usefulness to society. Further information about this fellowship is available from the Acoustical Society of America, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org.

Third Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan held in Honolulu, Hawaii

The Third Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan was held 2–6 December 1996 at the Sheraton Waikiki Hotel in Honolulu, Hawaii. The previous joint meetings were held in the Fall of 1978 and the Fall of 1988.

This meeting drew a total of 1652 registrants, which slightly exceeds the previous high at the meeting in Cambridge, Massachusetts in June 1994. The principal components of the total included 995 from North America (U.S.A.—944, Canada—46, and Mexico—5) and 477 from Japan. Attesting

to the international ties of our organizations, 179 of the registrants were from countries other than North America and Japan. There were 24 registrants from France, 21 from Sweden, 20 from England, 18 from Australia, 14 from Germany, 13 each from Korea and Denmark, 9 from The Netherlands, 6 each from Russia and New Zealand, 5 from Norway, 4 each from Italy and Singapore, 3 each from Belgium, Finland, and the People's Republic of China, 2 each from Hong Kong, Israel, and Ukraine, and 1 each from Ecuador, Greece, Peru, Portugal, Saudi Arabia, Scotland, and Taiwan.

There were 250 nonmembers in attendance of whom over 66% joined the Society at the time of registration. Additionally, 280 students were in attendance.

The 1335 papers, which far exceeded that at any previous meeting, were organized into 91 sessions and covered the areas of interest of all 13 Technical Committees and Technical Groups. The meeting was opened with a Plenary Session on Monday morning at which the Presidents of the Acoustical Society of America and the Acoustical Society of Japan welcomed all the registrants and introduced the two keynote speakers. A special congratulatory note was made of the 60th Anniversary of the Acoustical Society of Japan by the ASA President. The respective Chairs of the ASA and ASJ organizing committees for the meeting added their welcoming remarks. A full program of parallel technical sessions continued through Friday afternoon with an outstanding mix of invited and contributed papers, some of which were presented as posters. The Chair of each session was shared by representatives of the ASA and the ASJ in the general spirit of cooperation and friendship that was projected throughout the meeting. A strong level of activity in the ASA standards endeavors was shown by the 15 meetings on different subjects in this area.

The traditional tutorial presentation was continued at the Hawaii meeting. Ben Sharp of Wyle Laboratories, Arlington, VA, lectured on "Sound transmission through structures—Protecting occupants from exterior noise."

A seven-course Chinese banquet was held on Thursday evening with over 600 in attendance. The banquet was followed by the presentation of ASA awards, Special Distinguished Service Certificates, Acoustical Society of Japan Medals of Special Merit, and the announcement of newly elected Fellows of the Society. The Presidents of the two sponsoring societies thanked the local and technical meeting organizers for their prodigious efforts. The evening was concluded with a lively musical program presented by the Royal Hawaiian Band as a special treat.

The ASA President, Stan Ehrlich, and the ASA Vice President, Pat Kuhl, presented two Society Awards, two Science Writing Awards, and announced the election of 29 new Fellows. The Trent-Crede Medal was presented to Preston W. Smith, Jr. "for pioneering contributions to Statistical Energy Analysis and structural-acoustical interaction" (Fig. 1). The Distinguished Service Citation was presented to John C. Burgess "for contributions to international cooperation through the organization of joint meetings with the Acoustical Society of Japan" (Fig. 2).

The ASA President and Vice President also presented the Science Writing Award in Acoustics for Journalists to A. Richard Immel for his

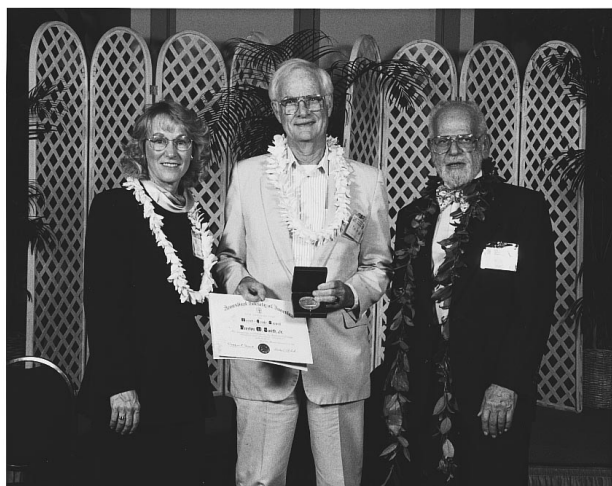


FIG. 1. Vice President Pat Kuhl (l) and President Stan Ehrlich (r) present the Trent-Crede Medal to Preston W. Smith, Jr.



FIG. 2. Vice President Pat Kuhl (l) and President Stan Ehrlich (r) present the Distinguished Service Citation to John C. Burgess.



FIG. 3. Vice President Pat Kuhl (l) and President Stan Ehrlich (r) present the Science Writing Award in Acoustics for Journalists to A. Richard Immel.



FIG. 4. Vice President Pat Kuhl (l) and President Stan Ehrlich (r) present the Science Writing Award for Professionals in Acoustics to Peter M. Narins.



FIG. 5. President Stan Ehrlich and Vice President Pat Kuhl with new Fellows of the Acoustical Society of America.

article, "Shh...those peculiar people are listening" published in the April 1995 issue of *Smithsonian* and the Science Writing Award for Professionals in Acoustics to Peter M. Narins for his article "Frog Communication" published in the August 1995 *Scientific American* (Figs. 3 and 4).

Election of the following persons to Fellow grade was announced: Sid P. Bacon, Avril Brenig, Catherine P. Browman, Antoine Chaigne, Ching-Sang Chiu, Raymond H. Dye, Earleen F. Elkins, Robert A. Fox, Sadaoki Furui, D. Felipe Gaitan, D. Wesley Grantham, Roger J. Hanson, George R. Harris, Mardi C. Hastings, David I. Havelock, Douglas H. Keefe, Sonoko Kuwano, Moises Levy, Richard L. McKinley, James B. Mehl, Elaine Moran, Seiichiro Namba, Victor Nedzelnitsky, Donna L. Neff, David A. Nelson, Robert V. Shannon, Timothy K. Stanton, Hideki Tachibana, and Richard L. Weaver (Fig. 5).

The President and the Vice President of the Acoustical Society of America presented Special Distinguished Service Certificates to the following distinguished members of the Acoustical Society of Japan for their special cooperation in the previous and the current joint meetings and their outstanding contributions to the friendship between the two societies: Juichi Igarashi, Sonoko Kuwano, Seiichiro Namba, and Hideki Tachibana (Fig. 6).

The President of the Acoustical Society of Japan, Hideki Tachibana, and the Vice President, Sadaoki Furui, presented Medals of Special Merit to the following distinguished members of the Acoustical Society of America for their outstanding contributions to the friendship between the two Societies: Robert E. Apfel, John C. Burgess, Stanley L. Ehrlich, Tony F. W. Embleton, Patricia K. Kuhl, and Jiri Tichy (Fig. 7).



FIG. 6. ASA Vice President Pat Kuhl (l) and ASA President Stan Ehrlich (r) present Special Distinguished Service Certificates to Juichi Igarashi, Sonoko Kuwano, Seiichiro Namba, and Hideki Tachibana.



FIG. 7. ASJ Vice President Sadaoki Furui (l) and ASJ President Hideki Tachibana (r) present Awards of Special Merit to Robert E. Apfel, John C. Burgess, Stanley L. Ehrlich, Tony F. W. Embleton, Patricia K. Kuhl, and Jiri Tichy.

The ASA President expressed the Society's thanks to the local committee for the excellent execution of the meeting, which clearly required meticulous planning. He introduced the Chair of the Meeting, John C. Burgess, who acknowledged the contributions of Tony F. W. Embleton, Technical Program Chair and members of the Hawaii Host Committee including: Whitlow W. L. Au, David L. Adams, Ronald A. Darby, Alexandra I. Tolstoy, Steve S. Black, Frances Burgess, Millicent Darby, Dorothy Au, and Maggie Adams.

John Burgess also thanked the ASA and ASJ members of the Technical Program Organizing Committee: Tony F. W. Embleton, Chairman for the ASA; Seiichiro Namba, Chairman for the ASJ; and Anthony A. Atchley, Associate Chairman.

For the ASA: Darrell R. Jackson and Alexandra I. Tolstoy, Acoustical Oceanography; Whitlow W. L. Au, Animal Bioacoustics; David L. Adams and Ronald A. Darby, Architectural Acoustics; Robin O. Cleveland, Bioreponse to Vibration/Biomedical Ultrasound; Stanley L. Ehrlich and George S. K. Wong, Engineering Acoustics; E. Carr Everbach, Musical Acoustics; Angelo J. Campanella and Joseph Pope, Noise; Anthony A. Atchley and E. Carr Everbach, Physical Acoustics; William A. Yost, Psychological and Physiological Acoustics; David I. Havelock, Signal Processing in Acoustics; Patricia K. Kuhl and Sigfrid D. Soli, Speech Communication; Gerard P. Carroll and J. Stuart Bolton, Structural Acoustics and Vibration; Stanley A. Chin-Bing and Guy V. Norton, Underwater Acoustics.

For the Acoustical Society of Japan: Hideki Tachibana, Architectural Acoustics; Hajime Miura, Engineering Acoustics; Seiichiro Namba, Musical Acoustics; Sonoko Kuwano, Noise; Kiyoshi Nakamura, Physical Acoustics; Toshio Sone, Psychological and Physiological Acoustics; Sadaoki Furui, Speech Communication; Yoshio Yamazaki, Signal Processing in Acoustics.

STANLEY L. EHRLICH
President 1996–1997

Short Courses

The Center for Professional Development at the University of Michigan in Ann Arbor offers a comprehensive program of short courses. These short courses have been designed to serve the education needs of engineers, scientists, and managers with experience in industry, government, and education. The courses will be held on the North Campus of the University of Michigan.

Following is a sampling of courses being offered:

| | |
|--------------|---|
| 16–18 June | Methods and Applications for Time Frequency Signal Analysis |
| 30 Jun–2 Jul | Automotive Sound Quality Engineering |

4–8 August Finite Elements in Mechanical and Structural Design:
Linear, Static Analysis
11–13 August Introduction to Numerical Acoustics

For a complete course schedule or for detailed information on courses contact J. Downs Herold, Director, Short Courses and Conferences, Center for Professional Development, Univ. of Michigan, College of Engineering, 273 Chrysler Center, 2121 Bonisteel Blvd., Ann Arbor, MI 48109-2092, Tel.: 313-647-7200; Fax: 313-647-7182.

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.

1997

6–8 June International Hearing Aid Conference IV, Iowa City, IA [Richard Tyler, Univ. of Iowa, Dept. of Otolaryngology—Head & Neck Surgery, 200 Hawkins Dr. C21-3GH, Iowa City, IA 52242-1078, Tel.: 319-356-2471; FAX: 319-353-7639; E-mail: rich-tyler@uiowa.edu].
15–20 June Eighth International Symposium on Nondestructive Characterization of Materials, Boulder, CO [Debbie Harris, The Johns Hopkins University, Ctr. for Nondestructive Evaluation, 102 Maryland Hall, 3400 N. Charles St., Baltimore, MD 21218, Tel.: 410-516-5397; FAX: 410-516-7249, E-mail: cnde@jhvmshcf.jhu.edu].
15–17 June NOISE-CON 97, State College, PA [Institute of Noise Control Engineering, P.O. Box 320, Arlington Branch, Poughkeepsie, NY 12603, Tel.: 914-891-1407; FAX: 914-463-0201].
16–20 June 133rd meeting of the Acoustical Society of America, State College, PA [ASA, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; FAX: 516-576-2377; E-mail: asa@aip.org, WWW: http://asa.aip.org].
9–13 July International Clarinet Association, Texas Tech Univ., Lubbock, TX [Keith Koons, Music Department, Univ. of Central Florida, P.O. Box 161354, Orlando, FL 23816-1354, Tel.: 407-823-5116; E-mail: kkoons@pegasus.cc.ucf.edu].

21–27 Aug.

7–11 Sept.

19–20 Sept.

22–24 Sept.

1–5 Dec.

22–26 June

13–17 Sept.

12–16 Oct.

1997 Conference on Implantable Auditory Prostheses, Pacific Grove, CA [Alena Wilson, Conference Coordinator, House Ear Inst., 2100 W. 3rd St., Los Angeles, CA 90057; Tel.: 213-353-7086; FAX: 213-413-0950; E-mail: alena@hei.org; WWW: http://www.rti.org/ciap97].

American Academy of Otolaryngology—Head and Neck Surgery, San Francisco, CA [American Academy of Otolaryngology—Head and Neck Surgery, One Prince St., Alexandria, VA 22314, Tel.: 703-836-4444; FAX: 703-683-5100].

Fifth Annual Conference on Management of the Tinnitus Patient, Iowa City, IA [Richard Tyler, Univ. of Iowa, Dept. of Otolaryngology—Head & Neck Surgery, 200 Hawkins Dr. C21-3GH, Iowa City, IA 52242-1078, Tel.: 319-356-2471; FAX: 319-353-7639; E-mail: rich-tyler@uiowa.edu].

Second Biennial Hearing Aid Research and Development Conference, Bethesda, MD [National Institute of Deafness and Other Communication Disorders, 301-970-3844; FAX: 301-907-9666; E-mail: hearingaid@tascon.com]. Deadline for abstracts is 15 March.

134th meeting of the Acoustical Society of America, San Diego, CA [ASA, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; FAX: 516-576-2377; E-mail: asa@aip.org, WWW: http://asa.aip.org].

135th meeting of the Acoustical Society of America/16th International Congress on Acoustics, Seattle, WA [ASA, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; FAX: 516-576-2377; E-mail: asa@aip.org, WWW: http://asa.aip.org].

American Academy of Otolaryngology—Head and Neck Surgery, San Francisco, CA [American Academy of Otolaryngology—Head and Neck Surgery, One Prince St., Alexandria, VA 22314, Tel.: 703-836-4444; FAX: 703-683-5100].

136th meeting of the Acoustical Society of America, Norfolk, VA [ASA, 500 Sunnyside Blvd., Woodbury, NY 11797, Tel.: 516-576-2360; FAX: 516-576-2377; E-mail: asa@aip.org, WWW: http://asa.aip.org].

Walter G. Mayer

Physics Department, Georgetown University, Washington, DC 20057

Austrian Acoustics Association founded

Several months ago a constituent assembly of 18 members had been formed to lay the groundwork for the formation of an Austrian Acoustics Association (AAA). On 7 December 1996 the assembly met to formally announce the foundation of the Oesterreichische Gesellschaft für Akustik OeGA and to elect the board of the new society. Eleven Technical Committees were established and their chairpersons were named. The first president of the association is E. Benes of the Technical University of Vienna; W. A. Deutsch of the Austrian Academy of Sciences was elected vice president.

Medals

The Institute of Acoustics (UK) has awarded the A. B. Wood Medal to Nicholas C. Makris of the Naval Research Laboratory in Washington, DC. He received the medal in recognition of his outstanding contributions to the science of underwater acoustics.

The French Acoustical Society has awarded the 1996 Médaille Etrangère to Herbert Überall of the Catholic University of America in Washington, DC. The medal was bestowed upon him in recognition of his contributions to the knowledge of acoustical resonance phenomena.

J. Edward White receives honor from Russian Academy

Dr. J. E. (Ed) White, a Fellow of the Society and professor emeritus in the Department of Geophysics at Colorado School of Mines was awarded the Russian Academy of Natural Sciences' Piotr L. Kapitsa Gold Medal in November 1996. One of the first American geophysicists to visit Russia, White established close ties with the Russian geophysical community. As a member of a six-man delegation sponsored by the U.S. Department of State in 1965, White obtained an overview of geophysical exploration in the Soviet Union and initiated correspondence with some of the Soviet geophysicists who were members of the return exchange delegation. In 1971, White displayed equipment at an exhibition in Moscow and took a post-convention tour to Akadengorodok. His proposal to spend an academic year in the Soviet Union was accepted as part of the exchange program between the U.S. National Academy of Sciences and the Soviet Academy of Sciences. While in Moscow at the Institute of Physics of the Earth, he was able to visit some earthquake observatories and centers of applied geophysical research in different parts of the Soviet Union. White visited the Soviet Union again in 1986 and 1992 to continue his scientific relations with the Russians.

Among his accomplishments, White is noted for his research on shear-wave prospecting, full-waveform acoustic logging, and attenuation of seismic waves. He has also authored several textbooks, one of which, *Underground Sound*, has been translated into Russian and Chinese. White holds 23 patents and is a former patent reviewer of this Journal. In 1989, he was elected to membership in the National Academy of Engineering.

Papers published in JASJ(E)

A listing of Invited Papers and Regular Papers appearing in the latest issue of the English language version of the *Journal of the Acoustical Society of Japan*, JASJ(E), was published for the first time in the January 1995 issue of the Journal. This listing is continued below.

The March issue of JASJ(E), Vol. 18, No. 2 (1997) contains the following papers:

- N. Tatsumoto, N. Kawano, and S. Fujii "Acoustic streaming by ultrasonic vibrating electrode (USVE)"
- S. Wei, M. Zhang, and K. Shimizu "Realization of over-easy characteristic of compressor/limiter on a digital signal processor"
- I. Nakayama "Voice timbre in autophonic production compared with that in extraphonic production"
- M. Akagi and T. Ienaga "Speaker individuality in fundamental frequency contours and its control"

M. Namba, H. Kamata, and Y. Ishida "An approach to speaker identification using DP-matched LVQ neural networks"

International Meetings Calendar

Below are announcements of meetings to be held abroad. Entries preceded by an * are new or updated listings with contact addresses given in parentheses. Month/year listings following other entries refer to issues of the *Journal* which contain full calendar listings or meeting announcements.

June 1997

- 3–5 **8th International Meeting on Low Frequency Noise & Vibration**, Gothenburg. 2/97
- 4–6 **International Conference on Computational Acoustics and its Environmental Applications**, Terni. 4/97
- 5–7 **Conference on ICP and Inner Ear Pressure**, Bath. 4/96
- 18–20 ***Larynx 97**, Marseille, France. (Secretariat Larynx 97, LPL, Université de Provence, 29 Av. Robert Schuman, 13621 Aix en Provence, France; Fax: +33 42 59 50 96; E-mail: larynx97@lpluniv-aix.fr; www.lpl.univ-aix.fr)
- 18–21 **3rd European Conference on Audiology**, Prague. 12/96
- 23–25 **1st International Conference on Marine Electromagnetics**, London. 2/97
- 24–27 **1st European Conference on Signal Analysis and Prediction**, Prague. 8/96
- 25–27 **12th Echocardiography Symposium and 9th Meeting of the International Cardiac Doppler Society**, Rotterdam. 6/96
- 25–27 **5th International Congress of the International Society of Applied Psycholinguistics (ISAPL '97)**, Porto. 10/96

July 1997

- 2–4 **Ultrasonics International '97**, Delft. 4/96
- 14–17 **6th International Conference on Recent Advances in Structural Dynamics**, Southampton. 2/97
- 21–25 **4th International Conference on Natural Physical Processes Related to Sea Surface Sound**, Chilworth Manor, Hampshire. 4/97

August 1997

- 16–19 ***Fechner Day '97**: Annual meeting of the International Society for Psychophysics, Poznan, Poland. (Web: www.ia.amu.edu.pl/fd97/)
- 18–22 **3rd EUROMECH Solid Mechanics Conference**, Stockholm. 10/96
- 19–22 **International Symposium on Musical Acoustics**, Edinburgh. 8/96
- 21–23 **ACTIVE 97 Inter-Noise Satellite Symposium**, Budapest. 6/96
- 24–27 **1997 World Congress on Ultrasonics**, Yokohama. 4/96
- 25–27 **Inter-Noise 97**, Budapest. 2/96
- 29–31 ***Pan-European Voice Conference**, Regensburg, Germany. (School of Logopedics, University of Regensburg, 93042 Regensburg, Germany; Fax: +49 941 953 1926)

September 1997

- 1–4 **Modal Analysis Conference (IMAC-XV Japan)**, Tokyo. 10/96
- 9–12 **31st International Acoustical Conference "Acoustics—High Tatra 97,"** High Tatra. 10/96

| | | | |
|----------------------|--|----------------------------|--|
| 10–12 | Biennial Conference New Zealand Acoustical Society , Christchurch. 4/96, 12/96 | March 1998 23–27 | DAGA 98 (German Acoustical Society Meeting) , Zürich. 8/96 |
| 10–12 | Biomechanics of Hearing , Stuttgart. 10/96 | | |
| 15–18 | 3rd EUROMECH Fluid Mechanics Conference , Göttingen. 10/96 | | |
| 16–19 | *44th Open Seminar on Acoustics , Jastrzebia Gora, Poland. (E. Kozaczka, Akademia Marynarki Wojennej, ul. Smidowicza 71, 81-919 Gdynia, Poland; Fax: +48 58 25 48 46; E-mail: amw@beta.nask.gda.pl) | May 1998 25–27 | Noise and Planning 98 , Naples. 2/97 |
| 18–19 | 4th Mexican Congress on Acoustics , Guanajuato. 4/96 | June 1998 8–10 | EAA/EEAA Symposium “Transport Noise and Vibration,” Tallinn. 10/96 |
| 18–20 | Intonation: Theory, Models and Applications , Athens. 2/97 | 9–12 | *8th International Conference on Hand-Arm Vibration , Umea, Sweden. (National Institute for Working Life, Physiology and Technology Department, P.O. Box 7654, 90713 Umea, Sweden; Fax: +46 90 165027; E-mail: hav98@niwl.se) |
| 22–25 | 5th European Conference on Speech Communication and Technology , Patras. 2/96 | 20–28 | *Joint Meeting of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America , Seattle, WA. (16th ICA Secretariat, Applied Physics Laboratory, 1013 NE 40th Street, Seattle, Washington 98105-6698, USA) |
| 23–26 | Fluid-structure Interaction in Acoustics , Delft. 10/96 | | |
| 27–28 | *Audio-Visual Speech Processing , Rhodes, Greece. (ESCA, ICP-Université Stendhal, BP 25X, 38040 Grenoble Cedex, France; E-mail: esca@icp.grenet.fr) | | |
| October 1997 | | November 1998 | |
| 7–10 | 1997 IEEE Ultrasonics Symposium , Toronto. 2/96 | 16–18 | Inter-Noise 98 , Christchurch. 4/96 |
| 8–10 | Acoustics Week in Canada 1997 , Windsor. 12/96 | 23–27 | ICBEN 98: Biological Effects of Noise , Sydney. 12/96 |
| 23–26 | Reproduced Sound 13 , Windermere. 2/97 | 30–4 | *5th International Conference on Spoken Language Processing , Sydney, Australia. (ICSLP Secretariat, Tour Hosts, GPO Box 128, Sydney NSW 2001, Australia; Fax: +61 2 9262 3135; E-mail: tourhosts@tourhosts.com.au; Web: http://cslab.anu.edu.au/icslp98) |
| 30–31 | *Swiss Acoustical Society Meeting , Bern, Switzerland. (Swiss Acoustical Society, P.O. Box 251, 8600 Dübendorf, Switzerland) | | |
| November 1997 | | March 1999 | |
| 19–21 | WESTPRAC'97 , Hong Kong. 10/96 | 15–19 | *Joint Meeting of EAA Forum Acusticum and 137th Meeting of the Acoustical Society of America , Berlin, Germany. (ASA, 500 Sunnyside Blvd., Woodbury, NY 11797, USA; Fax: +1 516 576 2377; E-mail: asa@aip.org) |
| 27–30 | *[New Date] IOA Autumn Conference: Environmental Noise , Windermere. 2/97 | | |
| December 1997 | | | |
| 15–18 | 5th International Congress on Sound and Vibration , Adelaide. 10/96 | | |
| February 1998 | | | |
| 2–6 | *Ultrasonic Technological Processes-98 , Moscow, Russia. (Secretariat UsTP-98, 64 Leningradski prospekt, MADI-TU, Moscow, Russia; Fax: +7 095 151 7911; E-mail: utp@madi.msk.su) | | |

REVIEWS OF ACOUSTICAL PATENTS

Daniel W. Martin

7349 Clough Pike, Cincinnati, Ohio 45244

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception Incorporated, Box 39536, Los Angeles, California 90039*

MAHLON D. BURKHARD, *31 Cloverdale Heights, Charles Town, West Virginia 25414*

SAMUEL F. LYBARGER, *101 Oakwood Road, McMurray, Pennsylvania 15317*

D. LLOYD RICE, *11222 Flatiron Drive, Lafayette, Colorado 80026*

CARL J. ROSENBERG, *Acentech Incorporated, 125 Cambridge Park Drive, Cambridge, Massachusetts 02140*

WILLIAM THOMPSON, JR., *The Pennsylvania State University, University Park, Pennsylvania 16802*

ERIC E. UNGAR, *Acentech Incorporated, 125 Cambridge Park Drive, Cambridge, Massachusetts 02140*

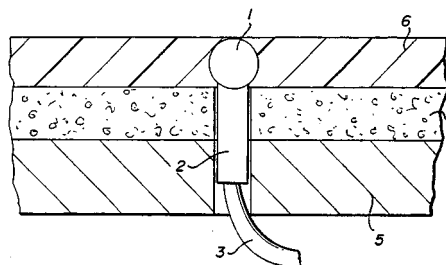
ROBERT C. WAAG, *University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642*

5,392,258

43.30.Xm UNDERWATER ACOUSTIC INTENSITY PROBE

T. B. Gabrielson *et al.*, assignors to the United States of America
21 February 1995 (Class 367/149); filed 12 October 1993

A neutrally buoyant underwater probe, for determining the true acoustic intensity via the simultaneous direct measurement of acoustic pressure and acoustic velocity, utilizes a moving-coil geophone embedded in a casting of syntactic foam plus a pair of hydrophones mounted on the exterior of that casting.—WT



5, called the acoustic shield, is constructed of a large rho-c material, such as stainless steel. It functions to shield the hydrophone from unwanted noise that might emanate from the interior of the vessel to which the whole assembly is attached. Layer 5 may, in fact, be the hull of the vessel. Layer 6 is a flow fairing made of rho-c material.—WT

5,469,404

43.30.Xm METHOD AND APPARATUS FOR SEISMIC EXPLORATION

H. P. Barker, Fulshear, TX and C. Lee, Houston, TX
21 November 1995 (Class 367/23); filed 11 August 1994

Eight seismic source air guns are arranged in a three-dimensional configuration and fired in time sequence in such a manner that the preponderance of usable output energy is concentrated in a narrow bandwidth at low frequencies.—WT

5,469,406

43.30.Yj METHOD OF EMITTING VERY LOW FREQUENCY HIGH POWER ACOUSTIC WAVES, AND CORRESPONDING TRANSDUCERS THEREFOR

A. A. Scarpitta *et al.*, assignors to L'Etat Francais
21 November 1995 (Class 367/171); filed in France 3 December 1993

A type V flextensional transducer comprises a number of electromechanical motor units mounted to the external surface of the flexible shell while the interior volume of the shell is free flooded except for a large body of low rho-c plastic foam that fills the main portion of that volume. This material provides pressure release action to the inward radiation.—WT

5,436,874

43.30.Yj METHOD AND APPARATUS FOR SENSING ACOUSTIC SIGNALS IN A LIQUID

P. M. Kuhn and F. P. Hodges, assignors to Martin Marietta Corporation
25 July 1995 (Class 367/176); filed 17 November 1993

By positioning a small omnidirectional hydrophone 1 at the surface of a composite layer structure where the two layers 4 and 5 have the appropriate acoustical impedance properties, a more or less hemispherical response pattern is realized. Layer 4, called the acoustic baffle, is a lossy material whose characteristic impedance is not much different from that of water so that sound waves impinging on that layer are significantly absorbed. Layer

5,449,964

43.38.Ar TRIGGERING TRANSDUCER APPARATUS FOR ACOUSTIC DEVICE

Michael J. Snyder, Tualatin, OR
12 September 1995 (Class 310/230); filed 23 May 1994

The triggering transducer is intended for use with musical instruments to reliably detect the onset of acoustic output from an instrument. The construction is somewhat reminiscent of phonograph pickups, employing a stylus-like member to contact the instrument and a piezoelectric transducing element.—MDB

5,440,630

43.38.Dv TELEPHONE RECEIVER WITH ELECTROSTATIC DISCHARGE PREVENTION

Takuro Yamaguchi *et al.*, assignors to Foster Electric, Ltd. and Mitel Corporation
8 August 1995 (Class 379/437); filed in Japan 8 February 1990

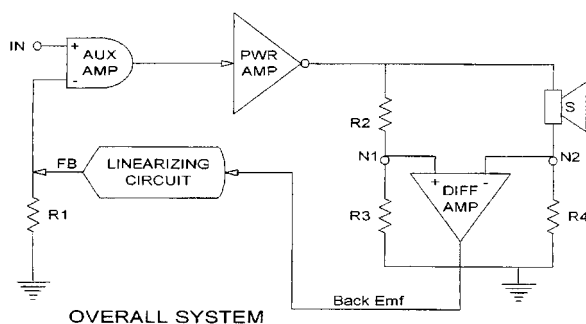
Electrostatic discharge between the user and an earphone can be annoying and may damage the earphone. Metallic shielding and insulation are placed within the structure to create a discharge path to a user that is significantly longer than internal paths to operationally safe components.—MDB

5,542,001

43.38.Ew SMART AMPLIFIER FOR LOUDSPEAKER MOTIONAL FEEDBACK DERIVED FROM LINEARIZATION OF A NONLINEAR MOTION RESPONSIVE SIGNAL

Martin Reiffin, Danville, CA
30 July 1996 (Class 381/96); filed 6 December 1994

This somewhat familiar diagram suggests that the novelty of the system must lie in its linearizing circuit. This includes two multiple-diode "correction generators" intended to cancel nonlinearities in positive and negative cone excursions. Whether the circuit can really be described as a



motional feedback system is arguable, but it should be able to reduce low-frequency distortion. The patent provides no actual performance data.—GLA

5,521,983

43.38.Ja SPEAKER SYSTEM FOR USE IN HIGH BACKGROUND NOISE ENVIRONMENTS

George W. Thompson III *et al.*, assignors to Vectra Corporation
28 May 1996 (Class 381/98); filed 28 October 1993

The patent illustrations show a typical public address horn design connected to a power amplifier through a bandpass filter centered at about 1 kHz. "The narrower range sound signal is better suited for penetrating high background noises and is more easily heard and understood by listeners in a high background noise environment when the sound signal frequency is significantly different than that of the background noise." Most audio engineers will agree with this since it has long been accepted practice.—GLA

5,536,984

43.38.Ja VOICE COIL ACTUATOR

Keith O. Stuart and Dennis C. Bulgatz, assignors to Aura Systems, Incorporated
16 July 1996 (Class 310/13); filed 7 June 1995

This is another variant of Aura Systems' radial magnet loudspeaker design. A more conventional magnet is included in the bottom of the gap to linearize the magnetic circuit.—GLA

5,539,262

43.38.Ja AXIALLY FOCUSED RADIAL MAGNET VOICE COIL ACTUATOR

Michael G. Strugach, assignor to Aura Systems, Incorporated
23 July 1996 (Class 310/13); filed 3 August 1994

Aura Systems' radial magnet loudspeaker design is tweaked with two additional washer-shaped magnets to focus the magnetic field in the gap and to minimize fringe effects.—GLA

5,546,470

43.38.Ja LOUDSPEAKER

Stefan Geisenberger, assignor to Nokia Technology GmbH
13 August 1996 (Class 381/197); filed in Germany 1 June 1994

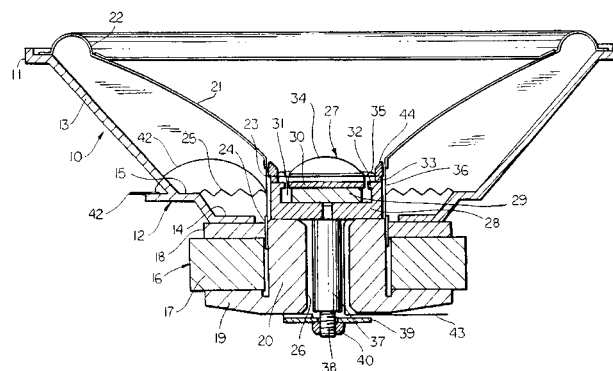
Instead of the usual centering spider, this interesting high-temperature loudspeaker employs a coil spring in conjunction with repulsive magnetic centering.—GLA

5,548,657

43.38.Ja COMPOUND LOUDSPEAKER DRIVE UNIT

Lawrence R. Fincham, assignor to KEF Audio (UK) Limited
20 August 1996 (Class 381/182); filed in the United Kingdom 9 May 1988

A patent can be granted almost nine years after its filing date. The geometry certainly looks familiar, but tweeter 27 employs a compact neodymium magnet so that its dome is almost as large as woofer voice coil tube



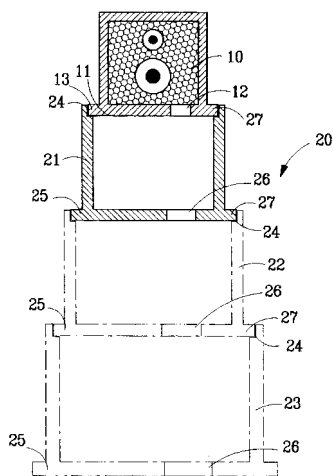
23 in which it nests. This allows the directional characteristics and acoustic origins of woofer and tweeter to be matched through the crossover range of frequencies.—GLA

5,550,926

43.38.Ja OVERLAP SOUND CASE

Ye-Ming Tsao, Taipei, Taiwan
27 August 1996 (Class 381/188); filed 8 September 1994

The individual sections of this loudspeaker enclosure can be sequentially disassembled, inverted, and nested together for transport. Air chambers communicate through vents 26, placed eccentrically to "... prevent sound wave resonance vibration." This reviewer has experimented with



similar multiple-chamber systems and can report that the performance may be quite different from that promised in the patent.—GLA

5,442,713

43.38.Kb MICROPHONE PACKAGING SCHEME

Bakulesh B. Patel and Rebecca C. McNally, assignors to Motorola, Incorporated
15 August 1995 (Class 381/168); filed 8 September 1992

The patent discloses a packaging scheme for a microphone in a communication device that is intended to reduce noise created by wind or breathing when close to the mouth of the user. A preferred microphone capsule is of the pressure gradient type. A porous membrane is mounted between the microphone and the openings on the opposing ends of the housing.—MDB

5,502,768

43.38.Lc REVERBERATOR

Kazuaki Shioda, assignor to Kawai Gakki Seisakusho
26 March 1996 (Class 381/61); filed in Japan 28 September 1992

This produces a reverberation effect for use with electronic musical instruments and can be implemented with analog or digital circuitry. It "...extracts only a specific frequency band signal from a distorted signal that is obtained by distorting an input tone signal, mixes the input tone signal with the specific frequency band signal, adds reverberation to the resultant signal, and generates a high frequency band signal to which optimal reverberation can be added."—GLA

5,550,925

43.38.Md SOUND PROCESSING DEVICE

Taizou Hori et al., assignors to Canon Kabushiki Kaisha
27 August 1996 (Class 381/98); filed in Japan 7 January 1991

Portable professional audio recorders may include a special low-cut filter for suppression of wind noise. The patent goes on at great length to explain why an operator may mistakenly switch in the filter when it is not needed, or fail to use the filter when it is needed. Moreover, a fixed filter is not the best way to cope with varying wind noise. The circuitry includes a variable filter, dynamically controlled by the low-frequency content of the microphone's output.—GLA

5,457,685

43.38.Si MULTI-SPEAKER CONFERENCING OVER NARROWBAND CHANNELS

Terrence G. Champion, assignor to the United States of America
10 October 1995 (Class 370/62); filed 15 July 1994

Of the patent's 30 pages, more than 20 are devoted to diagrams and computer source code. The system is for digital conferencing over narrow-band channels, "... which allows use by multiple simultaneous speakers and which allows for interfacing between systems which operate at different bit rates."—GLA

5,535,433

43.38.Si HANDS-FREE CIRCUITRY PROVIDING AUDIO SIGNAL ADJUSTMENT BASED ON SPEAKER VOLUME LEVEL

Osamu Kurokawa and Buntaro Sawa, assignors to Kabushiki Kaisha Toshiba
9 July 1996 (Class 455/79); filed in Japan 11 June 1991

Most speakerphone circuits introduce fixed attenuation in send and receive modes to provide semi-duplex operation, yet maintain a sufficient gain margin to avoid howling. This circuit is a more user-friendly design which compensates for the setting of the manual volume control.—GLA

5,574,795

43.38.Si HEADPHONE APPARATUS

Hideki Seki, assignor to Sony Corporation
12 November 1996 (Class 381/183); filed in Japan 31 July 1993

A double headset with a headband is shown. A sheathed spring element connected to the ends of the headband just above the earphones pulls them more firmly against the ears.—SFL

5,579,400

43.38.Si HEADPHONES

Burkhard Ballein, Wuppertal, Germany
26 November 1996 (Class 381/183); filed in Germany 17 July 1992

The patent shows two headphones mounted at the ends of adjustable-length eyeglass side pieces. Foam pads surrounding the headphones fit into the auricles of the user.—SFL

5,581,622

43.38.Si HEADSET

Kazuo Sakurai, assignor to Yashima Electric Company
3 December 1996 (Class 381/183); filed in Japan 29 August 1991

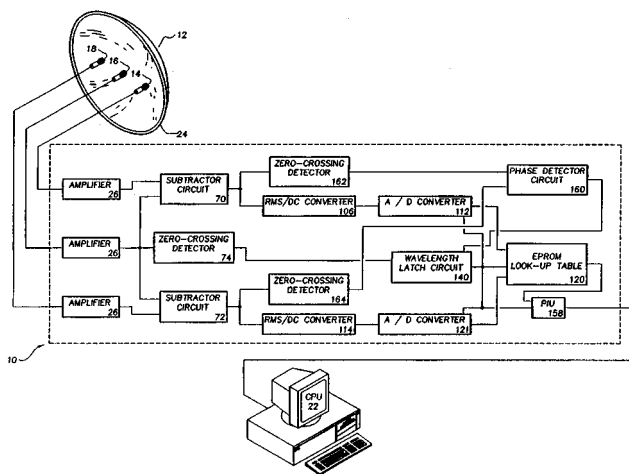
The headset comprises an auricle-insertion-type headphone, a microphone with boom, and connecting cords. The design is stated to improve wearing comfort.—SFL

5,352,364

43.38.Tj SYSTEM AND METHOD FOR MONITORING WILDLIFE

Douglas M. Bonham, Mountlake Ter., WA
19 September 1995 (Class 381/92); filed 7 December 1993

The signals received by three or more microphones **14, 16, 18** positioned relative to a parabolic reflector **12** enter an audio signal processing



circuit 10. Signals produced by circuit 10 are supplied to a central processing unit 22 to locate the wildlife producing a sound.—MDB

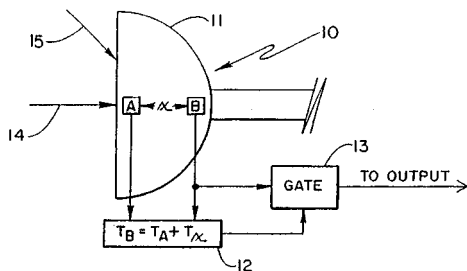
5,452,363

43.38.Tj DIRECTION SENSING MICROPHONE SYSTEM FOR TIME DIFFERENTIAL

Lynn J. Mader, Bismarck, ND

19 September 1995 (Class 381/92); filed 12 October 1993

The object of the microphone is to restrict the output of a sound reinforcement system to sounds originating at a particular location. "A pair of conventional microphone elements are spaced-apart within the microphone,



and the time it takes for the sound to pass between the two elements is measured and, if within a prescribed time, then the microphone is 'turned on' so that sound is permitted to reach or be gated to the output."—MDB

5,390,214

43.38.WI DIGITAL AUDIO BROADCASTING SYSTEM

John W. Hopkins, Norfolk VA and Anthony J. Impastato, Virginia Beach, VA

14 February 1995 (Class 375/37); filed 8 January 1993

The inventors envision a digital FM broadcasting system which includes complementary compression/expansion, similar to the processed analog recordings once offered by DBX and CBS. Also included is a receiver which retransmits to nearby standard FM tuners.—GLA

5,587,541

43.40.Cw MUSICAL INSTRUMENT STRINGS

Bruce M. McIntosh *et al.*, assignors to Zyex Limited

24 December 1996 (Class 84/297); filed in Great Britain 18 July 1995

This patent describes a musical instrument string material (for bowed or plucked strings) said to have significantly lower internal damping than previous strings, lower inharmonicity, and less variation caused by changes

in humidity. The material is a thermoplastic aromatic polyetherketone. The patent text quotes Lord Rayleigh's book and a *Journal* article by J. C. Schelling.—DWM

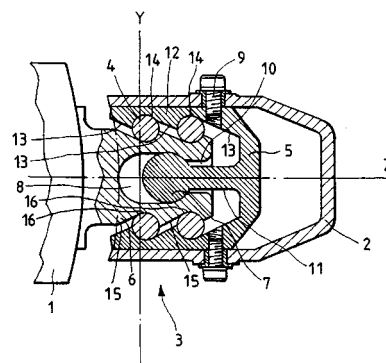
5,522,466

43.40.Tm VIBRATION DAMPING STRUCTURE FOR ELECTRIC HAMMER

Mutsuo Harada and Akihisa Yahagi, assignors to Hitachi Koki Company

4 June 1996 (Class 173/162.2); filed in Japan 28 October 1994

This patent relates to a means for isolating the handle of an electric hammer from vibrations caused by the hammer. Isolation is provided by resilient elements, with a cam-action arrangement employed to lock the handle and hammer body together more rigidly to facilitate withdrawing the tool when it gets stuck in the structure on which it is working. The handle 2 is connected to the tool body 1 via two or more arrangements like those shown in the figure. Each such arrangement consists of an elastic (plastic or



rubber) member 5 that has an essentially E-shaped cross section and is attached to the handle, an essentially C-shaped elastic member 6 that is attached to the tool body, and rollers 4. When the handle is pulled away from the tool body, as is done when the tool is withdrawn, the rollers and the wedge configuration of the surfaces of the elastic elements cause these elements to be locked together relatively rigidly. When the handle is pushed toward the tool body, as is done when the tool is being used, this wedging action is released, causing the connection between the handle and tool body to be relatively resilient.—EEU

5,528,005

43.40.Tm OSCILLATION ABSORBER FOR THE ABSORPTION OF STRUCTURE-BORNE SOUND

Oskar Bschorr and Klaus Zimmerman, assignors to Deutsche Aerospace AG

18 June 1996 (Class 181/208); filed in Germany 30 October 1993

This patent relates to waveguide absorbers that extract high-frequency vibrations (structure-borne sound) from structures. Waveguide absorbers discussed in the technical literature in effect are structural acoustical horns made, for example, in the form of extended tapered beams, that are intended to remove vibrational energy from vibrating structures and to dissipate this energy. The present patent presents several ways in which the aforementioned extended continuous horns can be replaced by arrays of spring-mass systems. One illustrative absorber consists of a stack of plates that are separated from each other by spacers which are located judiciously so that a motion normal to the plane of the plates causes all plates to flex, with plates that are further from the force application point deflecting more readily and with the farthest plate incorporating a damping layer.—EEU

5,533,422

43.40.Tm SYSTEM FOR ABSORBING TORSIONAL AND/OR BENDING VIBRATIONS

Frank H. Speckhart, Knoxville, TN

9 July 1996 (Class 74/573 R); filed 28 November 1994

The system described in this patent is a torsional vibration absorber (or "tuned damper") that is intended to suppress periodic forced vibrations of the shaft of a rotating machine. The system described here works on the centrifugal pendulum principle. However, instead of a classical pendulum that consists of a mass which pivots about a support, the system of this patent uses a rolling element mass that is free to roll inside a cavity with appropriately shaped interior walls. In a typical design several such cavities and rolling elements are built into a flywheel-like disk at uniform angular intervals. In one illustrated arrangement the cavities and rolling elements are cylindrical and are built into the disk near its circumference with their axes in the plane of the disk, so that the system can attenuate bending vibrations of the shaft which cause the disk to be deflected out of its plane. In another arrangement the cavities and rolling elements are spherical, so that the system can attenuate torsional, as well as bending vibrations.—EEU

5,538,373

43.40.Tm MACHINE TOOL VIBRATION ISOLATION SYSTEMEdward E. Kirkham, assignor to Giddings & Lewis, Incorporated
23 July 1996 (Class 409/131); filed 2 September 1994

Most versatile machine tools, such as the so-called machining centers, need to be rigidly connected to the supporting floor, so that alignment of the working tool relative to the workpiece is maintained. This rigid support requirement does not permit vibration isolation to be used between the machine structure and the floor. This patent deals with machine tool configurations that are stiff enough to maintain the required tool/workpiece alignment while permitting the machine tool to be supported from the floor via vibration isolators. In the configurations described in this patent the machine tool consists of two platforms, where one platform supports the workpiece and the other supports the working tool. The two platforms are interconnected by a number of legs that can move one platform relative to the other via a control system (but without active vibration control), enabling the tool to carry out the desired machining.—EEU

5,549,271

43.40.Tm VIBRATION CANCELER

Arno Hamaekers, assignor to Firma Carl Freudenberg

27 August 1996 (Class 248/621); filed in Germany 26 August 1993

This patent relates to a comparatively easily manufactured and assembled configuration of a dynamic absorber or "tuned damper." The absorber's spring element is made of a flexible elastomeric material and includes a metallic insert that locks the element into an appropriately shaped cavity in the absorber's inertial mass.—EEU

5,529,295

43.40.Vn VIBRATION-COMPENSATING MOUNT ASSEMBLY

Markus Leibach and Georg Feurer, assignors to Firma Carl Freudenberg

23 June 1996 (Class 26/140.15); filed in Germany 24 November 1993

This patent describes an active system which acts on an item that is isolated from a vibration source via a conventional mount. The active system here consists of a mass that is connected to the isolated item via a spring, and that is electro-dynamically actuated so as to oppose that item's vibratory motion, with the mass' motion confined along the axis of the mount.—EEU

5,442,710

43.40.Yq BODY-FELT SOUND UNIT AND VIBRATION TRANSMITTING METHOD THEREFOR

Akira Komatsu, assignor to Bodysonic Kabushiki Kaisha

15 August 1995 (Class 381/24); filed in Japan 30 November 1990

Claim 1 is quoted in part: "A body-felt vibration unit having a vibration transmitting member imbedded in a human body support member, and an electro-mechanical vibration transducer attached to a vibration receiving plate fixed to said vibration transmitting member... thereby transmitting varying vibration to a human body on the body support member, said electro-mechanical vibration transducer and said vibration receiving plate being disposed substantially perpendicular to said vibration transmitting member, wherein vibration generated from said transducer is transmitted along an extending direction of said vibration receiving plate, thereby transmitting transverse vibration"—MDB

5,452,612

43.40.Yq MULTI-MODE ACCELEROMETER

Peter R. Smith and John C. Henry, assignors to The Whitaker Corporation

26 September 1995 (Class 73/514.34); filed 23 December 1991

The accelerometer has a piezoelectric, preferably polymer, transducing element. It is in the form of an elongated continuous flat and narrow cross-section beam. Output is generated in response to deflection of the free ends by shock in either or both linear and torsion modes. Less need for matched transducer elements is a suggested feature of the construction.—MDB

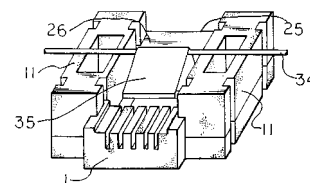
5,551,198

43.50.Gf SOUND COLLECTING BLOCK AND SOUND ABSORBING WALL SYSTEM

Cecil F. Schaaf and Craig R. Schaaf, Flint, MI

3 September 1996 (Class 52/604); filed 9 May 1995

The patent discloses a series of specially configured blocks which together form a sound barrier wall or retaining wall. The system of interlocking blocks is quite deep, and the front face of some of the blocks has



vertical slots to allow sound penetration to the cavity, where glass fiber in the cavity will absorb sound. The barrier wall could be erected without mortar, and would be used along a highway.—CJR

5,579,614

43.55.Ev ACOUSTICAL SYSTEM, A PART THEREFOR AND METHOD OF MAKING SAME

Gordon J. Dorn, DeKalb, IL

3 December 1996 (Class 52/144); filed 2 January 1996

This is just a vertically split cone assembly which purports to improve the acoustics in a room by reducing resonance and focusing, and by promoting sound diffusion. The cone is split from its vertex to a diameter of its base, along its vertical axis. The cones are arranged top to bottom to form a sound diffusing plane.—CJR

5,544,249

43.55.Lb METHOD OF SIMULATING A ROOM AND/OR SOUND IMPRESSION

Martin Opitz, assignor to AKG Akustische U. Kino-Geräte GmbH
6 August 1996 (Class 381/63); filed in the United Kingdom 26 August 1993

This is a clever method of simulating desired acoustical characteristics in real time with comparatively little computational overhead. The system employs a kind of perceptual encoding: "it is merely necessary to know those portions of the room impulse responses which in accordance with a critical selection are essential for the auditory impression." Readers interested in this subject will want to study this patent.—GLA

5,557,078

43.55.Ti ACOUSTICAL BARRIER

Matthew J. Holwerda, assignor to Cascade Engineering, Incorporated
17 September 1996 (Class 181/208); filed 14 September 1994

This acoustical barrier is for use in a vehicle to reduce or eliminate the transfer of noise from the engine to the passenger compartment. The barrier comprises an adaptable metal layer and a mass and foam layer intended to seal tightly around penetrations for cables and such.—CJR

5,577,507

43.58.Ls COMPOUND LENS FOR ULTRASOUND TRANSDUCER PROBE

Jonathan E. Snyder *et al.*, assignors to General Electric Company
26 November 1996 (Class 128/663.01); filed 21 November 1994

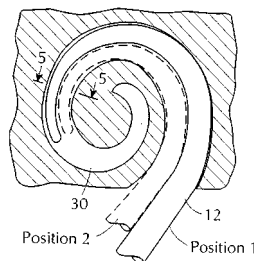
This compound lens has inner and outer parts. The inner part of the lens has a convex cylindrical front face and a rear face that is acoustically coupled to the front face of the transducer array. The outer part of the lens has a concave cylindrical rear face that is acoustically coupled to the convex cylindrical front face of the inner part of the lens. While the inner part of the lens may be a conventional silicone rubber material, the outer part of the lens is made of a medium such as polymethylpentene, nylon, or high-density polyethylene that has a higher acoustic velocity as well as greater durability and chemical resistance than silicone rubber.—RCW

5,578,084

43.66.Ts SELF-CURVING COCHLEAR ELECTRODE ARRAY

Janusz Kuzma *et al.*, assignors to Cochlear, Ltd. and University of Melbourne
26 November 1996 (Class 623/10); filed in Australia 27 September 1991

The electrode array for a cochlear implant is formed in two layers. The first layer is formed from a bio-compatible material and contains electrode contacts and connecting leads. The second layer is formed from a biocompatible material having a property of a controlled rate of expansion when



exposed to the water contained in body fluids. This material expands in use and curves the implanted array to enable more effective stimulation.—SFL

5,586,188

43.66.Ts HEARING AID

Peter Renggli *et al.*, assignors to Ascom Audiosys AG
17 December 1996 (Class 381/69.2); filed in Switzerland 10 May 1994

The patent describes a connecting device of small size that is used for attaching a programming cable to a programmable hearing aid.—SFL

5,573,015

43.66.Vt EXTRUDED EAR PLUG

Colin D. Williams, Charlotte, NC
12 November 1996 (Class 128/864); filed 28 March 1995

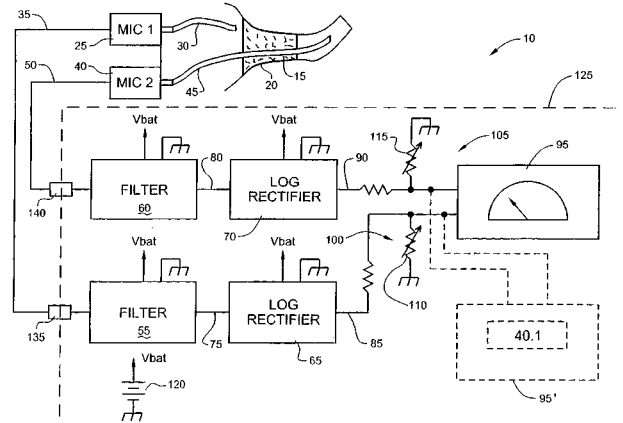
The patent shows a cylindrical ear protector made mostly of an outer compressible material, but having a more rigid central core. Means for extruding the plugs are also shown.—SFL

5,577,511

43.66.Yw OCCLUSION METER AND ASSOCIATED METHOD FOR MEASURING THE OCCLUSION OF AN OCCLUDING OBJECT IN THE EAR CANAL OF A SUBJECT

Mead C. Killion, assignor to Etymotic Research, Incorporated
26 November 1996 (Class 128/746); filed 29 March 1995

The patent shows an instrument for measuring the degree of occlusion of an object, such as an earmold, in the ear canal. Two microphones with tube inputs, one outside the occluding element and one inside the ear canal



beyond the occlusion, feed signals into logarithmic amplifiers whose rectified outputs are added to determine the difference between outside and inside sound-pressure levels.—SFL

5,550,924

43.72.Ew REDUCTION OF BACKGROUND NOISE FOR SPEECH ENHANCEMENT

Brant M. Helf and Peter L. Chu, assignors to PictureTel Corporation
27 August 1996 (Class 381/94); filed 7 July 1993

This speech enhancer processes the signal to reduce the perceptual effect of constant, additive noise which was unintentionally recorded with the speech signal. The input signal is filtered through one or more narrow-band notches which are tuned to particular frequencies of the noise signal. The result is transformed via a FFT, and several measures are taken to estimate signal magnitude, stationary components, background components,

and a running minimum. Based on these measures, a spectral gain vector is determined. The gains are subjected to spectral and temporal critical band spreading, then applied to the speech data as it is inverse transformed back to an output waveform.—DLR

5,553,190

43.72.Gy SPEECH SIGNAL TRANSMISSION METHOD PROVIDING FOR CONTROL

Tomoyuki Ohya *et al.*, assignors to NTT Mobile Communications Network

3 September 1996 (Class 395/2.1); filed in Japan 28 October 1991

Several techniques are described for improving the performance of a voice operated (VOX) speech coding system. Earlier VOX systems have typically sent an initializing preamble when voice activity is first detected. If the receiver misses the preamble due to transmission errors, an unnatural sound results as subsequent packets are decoded as speech without proper initialization. This system effectively inserts a preamble when required to restart the speech decoder. During nonspeech intervals, the receiver will insert a precoded noise sound to avoid an abrupt transition to complete silence.—DLR

5,553,191

43.72.Gy DOUBLE MODE LONG TERM PREDICTION IN SPEECH CODING

Tor B. Minde, assignor to Telefonaktiebolaget LM Ericsson

3 September 1996 (Class 395/228); filed in Sweden 27 January 1992

This variation on a code-excited LP (CELP) vocoder uses a new method of processing the long-term (pitch) prediction. The method combines open loop and closed loop tracking of the long-term delay values and the corresponding search for the optimum gain and excitation vector. A high-resolution (noninteger) delay value allows more accurate pitch tracking.—DLR

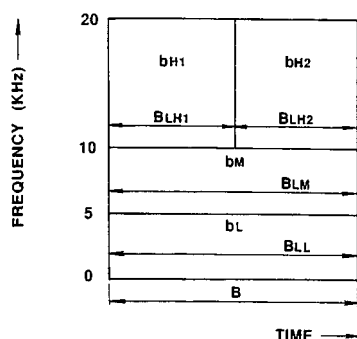
5,553,193

43.72.Gy BIT ALLOCATION METHOD AND DEVICE FOR DIGITAL AUDIO SIGNALS USING AURAL CHARACTERISTICS AND SIGNAL INTENSITIES

Kenzo Akagiri, assignor to Sony Corporation

3 September 1996 (Class 395/2.38); filed in Japan 7 May 1992

This speech coder system allocates quantization bits separately to individual sub-blocks of the time-frequency analysis space according to masking and equal-loudness criteria. Quadrature mirror (QMF) band-splitting filters produce octave-interval filter bands which are then analyzed by



modified discrete cosine transform (MDCT). The total bit allocation is divided into two parts, one part based on the critical band hearing criteria and a second part related to the signal energy. After both bit allocation strategies are performed, the partial bit allocations are summed in each of the time-frequency blocks.—DLR

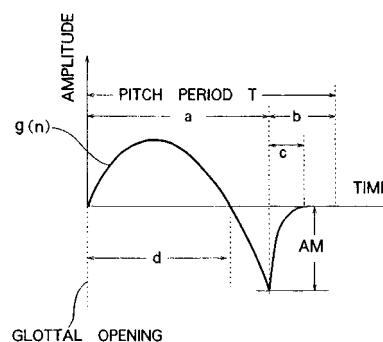
5,553,194

43.72.Gy CODE-BOOK DRIVEN VOCODER DEVICE WITH VOICE SOURCE GENERATOR

Katsushi Seza *et al.*, assignors to Mitsubishi Denki Kabushiki Kaisha

3 September 1996 (Class 395/2.3); filed in Japan 25 September 1991

This vocoder performs separate spectrum modeling and codebook searches for both the predictive (auto-regressive or AR) and the all-zero (moving average or MA) aspects of the speech analysis. It should therefore



be called an ARMA vocoder. The codebook search synthesis loop includes both AR and MA codebooks, as well as an additional voice source signal. The voice source generator produces a signal which corresponds to the derivative of the glottal airflow waveform.—DLR

5,550,893

43.72.Kb SPEECH COMPENSATION IN DUAL-MODE TELEPHONE

Alireza R. Heidari, assignor to Nokia Mobile Phones Limited

27 August 1996 (Class 379/59); filed 31 January 1995

This cellular telephone circuit contains two independent speech transmission systems, either of which may be selected to match the coding method of the base station. However, a change in voice quality is perceived when the speech coding system switches from PCM format (referred to here as "analog") to the compress/expand coding ("digital") format. In order to balance the speech quality between the two modes, an internal back-to-back compress/expand process is performed on the speech samples as they are processed in "analog" mode.—DLR

5,553,120

43.72.Kb TELEPHONIC-INTERFACE GAME CONTROL SYSTEM

Ronald A. Katz, Los Angeles, CA

3 September 1996 (Class 379/88); filed 10 July 1985

This is a telephone interface control system which allows many individuals to participate in a game format such as a call-in lottery game. Callers would use the standard DTMF telephone keypad to enter plays or navigate the system, and would receive audio prompts through the telephone handset. A caller ID unit would provide the system with rapid access to demographics or other caller-related information.—DLR

5,553,121

43.72.Kb VOICE RESPONSE SYSTEM

David G. Martin and Lawrence L. Porter, assignors to IBM Corporation

3 September 1996 (Class 379/88); filed in the United Kingdom 19 August 1994

This voice response system records and averages the user's response times and uses the results to adjust the verbosity or complexity or other aspects of the prompts. For example, a sequence of very rapid user re-

sponses might cause the system to select an abbreviated form of the query prompts. A user's response time averages are saved between sessions so that a new "getting acquainted" sequence will not be required for each repeat session.—DLR

5,546,499

43.72.Ne SPEECH RECOGNITION SYSTEM UTILIZING PRE-CALCULATED SIMILARITY MEASUREMENTS

Thomas E. Lynch *et al.*, assignors to Kurzweil Applied Intelligence, Incorporated
13 August 1996 (Class 395/2.49); filed 27 May 1994

Earlier patents have presented the idea of replacing the input speech spectral analysis vectors with some sort of standardized feature vectors, often discovered by clustering or other training method. This allows the use of precomputed distance measures, which, in principle, could greatly speed up the search process. In practice, however, the resulting distance metric table would be enormous. This patent takes that approach, but reduces the memory requirement by classifying the distance values, storing only those giving the closest matches. The resulting table is stored using sparse matrix techniques.—DLR

5,546,500

43.72.Ne ARRANGEMENT FOR INCREASING THE COMPREHENSION OF SPEECH WHEN TRANSLATING SPEECH FROM A FIRST LANGUAGE TO A SECOND LANGUAGE

Bertil Lyberg, assignor to Telia AB
13 August 1996 (Class 395/2.86); filed in Sweden 10 May 1993

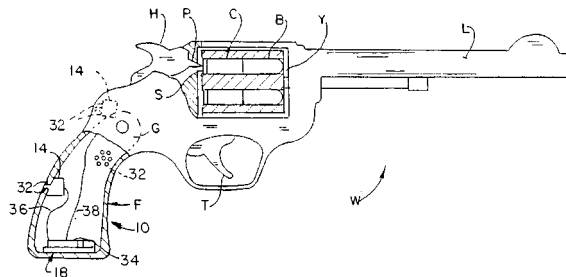
Although speech synthesis efforts have long included attempts to synthesize realistic intonation contours, prosodic analysis has not previously been used in an automatic language translator system. The recognition unit includes mechanisms for intonation and duration analysis and tables of language characteristics for the input language. A text-to-speech synthesizer contains language information tables for the target language. No further details are given, leaving open all of the real questions, such as how the language tables would work and how the recognized prosodic patterns might be mapped to another language.—DLR

5,546,690

43.72.Pf AUDIO CONTROLLED GUN LOCKING MECHANISM

Gary Ciluffo, Vincennes, IN
20 August 1996 (Class 42/70.11); filed 4 January 1995

This gun locking mechanism would use a voice analyzer built into the firearm to control the operation of a safety lock which would prevent firing the weapon except on receipt of the owner's voice pattern. As described, the device would trigger on "audio frequency signals such as a word or words"



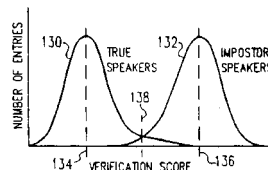
which are processed by a speech recognition chip. There is no discussion of the processing techniques used, the types of speaker verification errors which could occur, or how the design would favor certain types of errors.—DLR

5,548,647

43.72.Pf FIXED TEXT SPEAKER VERIFICATION METHOD AND APPARATUS

Jayant M. Naik and George R. Doddington, assignors to Texas Instruments, Incorporated
20 August 1996 (Class 381/42); filed 3 April 1987

This patent presents several components of a strategy to improve performance of a speaker verification system. The candidate user speaks an ID code, a two-word phrase, and a random-word phrase. Following LPC analysis and filter bank band separation, principal components analysis is performed and the results are dynamic time warped to match reference tem-



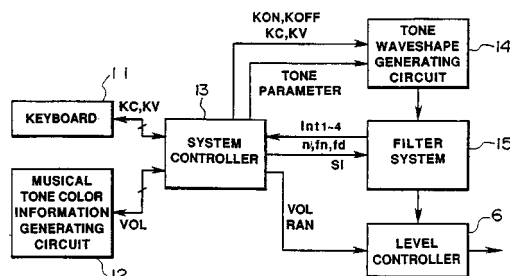
plates. If a retest is required, a new random-word phrase will be prompted and acceptance criteria may be altered depending on the true user's history of percentage of failures.—DLR

5,583,309

43.75.Tv FILTER APPARATUS FOR AN ELECTRONIC MUSICAL INSTRUMENT

Yoshio Fujita, assignor to Yamaha Corporation
10 December 1996 (Class 84/622); filed in Japan 4 October 1989

"A musical tone generating apparatus ... consists of a musical tone source, a coefficient generator, a digital filter and so on. The coefficient generator generates at least one time dependence coefficient, and applies it to output of the digital filter. The parameter controller changes at least one parameter, which designates tone color of the musical tone, in accordance with the time dependent coefficient to thereby realize time-variant changing rate of envelope of musical tone. And the coefficient generator and the



parameter controller are used by means of time sharing technique. That is, the coefficient is used as plural units, wherein each unit generates a coefficient in each stage, and the digital filter is also used as plural units. Each digital filter unit changes a parameter of tone color of the musical tone in accordance with the coefficient supplied thereto."—DWM

5,589,654

43.75.Tv ELECTRONIC DANCE FLOOR SYSTEM

Kern T. Konwiser and Kip M. Konwiser, Marina Del Rey, CA
31 December 1996 (Class 84/600); filed 7 March 1996

This dance floor system is assembled from sections of panels, each of which operates an electrical switch when stepped upon. Each switch is connected through a Musical Instrument Digital Interface (MIDI) converter to a MIDI equipped sound source. During dance routines the dancer steps or taps on selected floor panels, triggering the converter to energize musical sounds programmed for each specific panel. For tap dancing, the tapping sounds are used to enhance the sounds controlled by the dance floor.—DWM

43.75.Wx MUSICAL TONE SYNTHESIS SYSTEM HAVING SHORTENED EXCITATION TABLE

**Julius O. Smith III, assignor to Leland Stanford Junior University
24 December 1996 (Class 84/659); filed 1 September 1994**

One musical tone synthesis method is to synthesize "in accordance with the mechanisms which occur in natural musical instruments." When this technique has been applied to the synthesis of string instrument tone it has required a complex filter simulating the musical instrument body. The present patent eliminates the need for this filter through the use of a filtered delay loop which is excited by an excitation signal which corresponds to a partial impulse response of the body filter.—DWM

43.80.Pe METHOD AND APPARATUS FOR MEASURING INTERNAL TISSUE CHARACTERISTICS IN FEED ANIMALS

William C. Pratt, assignor to Micro Chemical, Incorporated
12 November 1996 (Class 128/660.07); filed 27 February 1995

This method and apparatus include an ultrasonic imaging unit and a unit that contains and dispenses a coupling liquid that facilitates ultrasonic transmission through the hide of an animal.—RCW

43.80.Qf DIAGNOSTIC ULTRASOUND APPARATUS

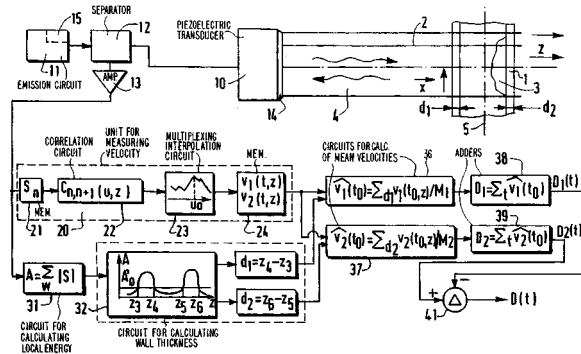
Mawaru Murashita *et al.*, assignors to Aloka Company
29 October 1996 (Class 128/660.04); filed in Japan 25 January
1994

This apparatus permits m -mode display of data along an arbitrarily oriented, user-selected line in a sequence of b -scan images.—RCW

43.80.Qf METHOD OF AND APPARATUS FOR THE DETECTION AND CHARACTERIZATION OF A SEGMENT OF AN ARTERY BY ULTRASONIC ECHOGRAPHY

Odile Bonnefus, assignor to U.S. Philips Corporation
3 December 1996 (Class 128/661.04); filed in France 25 March
1994

In this method and apparatus, time-domain correlation–interpolation signal processing is used to determine the radial displacement of artery walls. The correlation–interpolation process is performed between colinear



excitation lines obtained a few milliseconds apart. This yields velocity and amplitude values that are subsequently employed to determine dilation and compression curves for the artery.—RCW

43.80.Sh HIGH-INTENSITY ULTRASOUND THERAPY METHOD AND APPARATUS WITH CONTROLLED CAVITATION EFFECT AND REDUCED SIDE LOBES

**Jean-Yves Chapelon, assignor to Technomed Medical Systems and
Institut National
12 November 1996 (Class 601/2); filed 28 February 1995**

The signal generator in this method and apparatus supplies a random or pseudorandom wideband signal to a therapeutic transducer. Such signals permit the use of lower peak intensities to deliver a given amount of energy. The wideband nature of the signals reduces secondary focusing phenomena behind periodic or quasiperiodic structures.—RCW

43.80.Vj METHOD FOR MOTION ENCODING OF TISSUE STRUCTURES IN ULTRASONIC IMAGING

Bjørn Olstad, assignor to Vingmed Sound A/S
29 October 1996 (Class 128/660.07); filed in Norway 4 October
1994

In this method, the motion of a structure within a selected region is determined by cross-correlating echos from the region at two instants of time and then visualizing the structure as well as its estimated motion.—RCW

43.80.Vj METHOD FOR COMBINING ULTRASOUND VECTOR DATA FROM MULTIPLE FIRINGS TO IMPROVE IMAGE QUALITY

**Thomas L. Deitrich and Rowland F. Saunders, assignors to
General Electric Company
29 October 1996 (Class 128/661.01); filed 23 November 1994**

This method uses a compounder and a splicer to combine vector data from multiple firings along a scan line to improve image quality. The compounder contains an accumulator that sums two or more vector data sets and then scales the result. The splicer joins compounded vector data to form a single output vector using a weight function to combine data from transmissions focused at different ranges.—RCW

43.80.Vj METHOD AND APPARATUS FOR REAL-TIME CONCURRENT ADAPTIVE FOCUSING IN AN ULTRASOUND BEAMFORMER IMAGING SYSTEM

**J. Nelson Wright *et al.*, assignors to Acuson Corporation
5 November 1996 (Class 128/661.01); filed 5 August 1994**

Adaptive focusing is achieved by measuring delay correction values that correspond to a region of aberration concurrently during scan line formation in b-scan imaging. The corrections are updated at the scan line rate and applied to correct the focus of all scan lines during transmission and reception regardless of the imaging mode, scan geometry, or center frequency.—RCW

5,570,693

43.80.Vj METHOD AND APPARATUS FOR ULTRASOUND IMAGING AND ATHERECTOMY

Yue-Teh Jang and Axel F. Briskin, assignors to Cardiovascular Imaging Systems, Incorporated
5 November 1996 (Class 128/660.06); filed 15 December 1994

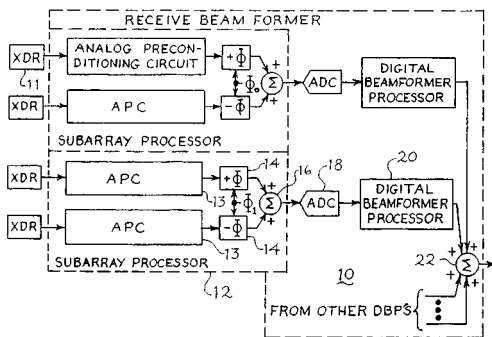
In this method and apparatus, an ultrasonic transducer in a catheter is fixed to a cutter. The transducer is moved longitudinally in a fixed radial position and ultrasonic reflections are processed to form an image of an axial plane determined by the radial position of the transducer. Other axial planes are imaged by rotating the transducer to a different angular orientation.—RCW

5,573,001

43.80.Vj ULTRASONIC RECEIVE BEAMFORMER WITH PHASED SUB-ARRAYS

Joseph G. Petrofsky *et al.*, assignors to Acuson Corporation
12 November 1996 (Class 128/661.01); filed 8 September 1995

This beamformer supplies receive echo signals to subarray processors 12, each of which includes at least one phase shifter 14 and summer 16. The phase angles for any one of the subarray processors sum approximately to zero and are independent of time delay in an associated digital beamformer



processor 20. The time resolution of delay in the digital beamformer processors is essentially as fine as the time resolution of the phase shifters. The phase shifter angles are updated at a slower rate than the focusing update rate of the beamformer processors.—RCW

5,574,212

43.80.Vj AUTOMATED SYSTEM AND METHOD FOR TESTING RESOLUTION OF ULTRASOUND SCANNERS

Ernest L. Madsen *et al.*, assignors to Wisconsin Alumni Research Foundation
12 November 1996 (Class 73/1 DV); filed 14 June 1995

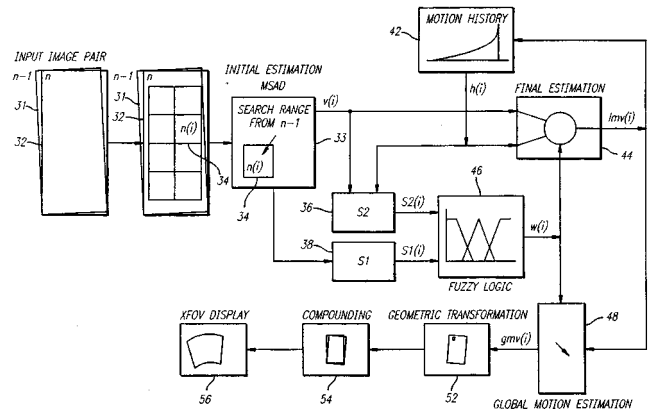
In this patent an ultrasonic phantom that contains background material mimicking the ultrasonic characteristics of tissue and coplanar spherical target lesions with an ultrasonic contrast relative to the background material is imaged using the ultrasonic scanner to be tested. Digitized images of background material and of focal lesions are obtained. A lesion signal-to-noise ratio is then calculated. This calculation employs pixels from a region that contains a lesion, pixels from a region that contains background echos, and a standard deviation of averaged pixel values in the background region. The proximal and distal depth range limits of detectability for a given lesion diameter and contrast are then determined from the number of pixels with a signal-to-noise ratio greater than a threshold.—RCW

5,575,286

43.80.Vj METHOD AND APPARATUS FOR GENERATING LARGE COMPOUND ULTRASOUND IMAGE

Lee Weng and Arun P. Tirumalai, assignors to Siemens Medical Systems, Incorporated
19 November 1996 (Class 128/653.1); filed 31 March 1995

A large compound ultrasonic image is produced by combining images obtained while the ultrasonic transducer of an imaging instrument is moved. Consecutive image frames are correlated to derive transducer motion infor-



mation. The correlation algorithm uses a fast adaptive coarse/fine minimum-sum-absolute-difference search strategy. Compound images are then formed using the estimates of transducer motion.—RCW

5,575,290

43.80.Vj COARSE-FINE ULTRASOUND TRANSDUCER ARRAY FOR MEDICAL IMAGING

Tat-Jin Teo and John R. Klepper, assignors to Siemens Medical Systems, Inc.
19 November 1996 (Class 128/661.10); filed 30 June 1995

This transducer array employs a different average row spacing of elements in elevation for reception and transmission. A relatively coarse row spacing for reception and a relatively fine row spacing for transmission lead to different transmission and reception beam patterns in elevation and permit the number of processing channels to be reduced or permit an increase in the aperture size without increasing the number of transducer elements or processing channels. The array includes dedicated transmit-only elements and/or dedicated receive-only elements as well as time-shared transmit/receive elements.—RCW

5,575,291

43.80.Vj ULTRASONIC COUPLER

Kenichi Hayakawa and Kiyoto Matsui, assignors to Fujitsu Ltd.
19 November 1996 (Class 128/662.03); filed in Japan 17 November 1993

This coupler intended for use between an ultrasonic probe and a subject to be imaged is made of a gel containing 3%–6% polyvinyl alcohol. The polyvinyl alcohol is at least 98.0 mole percent in the degree of saponification and has a hardness in the range of 0.5×10^4 – 2.0×10^4 dyne/cm². The coupler is molded with a shape that fits on the ultrasonic probe with which the coupler is used.—RCW

Speech perception as pattern recognition^{a)}

Terrance M. Nearey

Department of Linguistics, University of Alberta, Edmonton, Alberta T6G 2E7, Canada

(Received 21 December 1995; accepted for publication 16 December 1996)

This work provides theoretical and empirical arguments in favor of an approach to phonetics that is called double-weak. It is so called because it assumes relatively weak constraints both on the articulatory gestures and on the auditory patterns that map phonological elements. This approach views speech production and perception as distinct but cooperative systems. Like the motor theory of speech perception, double-weak theory accepts that phonological units are modified by context in ways that are important to perception. It further agrees that many aspects of such context dependency have their origin in natural articulatory processes. However, double-weak theory sides with proponents of auditory theories of phonetics by accepting that the real-time objects of perception are well-defined auditory patterns. Because speakers find ways to obey “orderly output conditions” (Sussman *et al.*, 1995), listeners are able to successfully decode speech using relatively simple pattern-recognition mechanisms. It is suggested that this situation has arisen through a stylization of gestural patterns to accommodate real-time limits of the perceptual system. Results from a new perceptual experiment, involving a four-dimensional stimulus continuum and a 10-category /hVC/ response set, are shown to be largely compatible with this framework. © 1997 Acoustical Society of America. [S0001-4966(97)05704-4]

PACS numbers: 43.10.Ln, 43.71.An, 43.71.Es [RAF]

INTRODUCTION

In this paper, I will sketch a framework called the *double-weak* theory of speech perception. Its main empirical hypothesis is that speech cues can be directly mapped onto phonological units of no larger than phoneme size. To put this theory in perspective, some underlying assumptions of current theories of phonetic specification will be examined. It is argued that most theories involve *a priori* constraints that may be too strong to stand the light of available data. The alternative advocated here requires only weaker versions of key assumptions of other theories. Many arguments presented below are summaries, elaborations, and minor revisions of views expressed elsewhere (Nearey, 1990, 1991, 1992, 1995). A summary of the theoretical and empirical work from the literature and results from a new experiment are presented. These generally support the main hypothesis of double-weak theory.

I. THEORETICAL BACKGROUND

Phonetics involves three logically distinct domains. The first of these is discrete and *symbolic*, consisting of one or more layers of phonological units (objects such as distinctive features, segments, moras, syllables, etc.). The other two domains are quasicontinuous and physical. They are the *gestural* (or articulatory) on the one hand and the *auditory* (or acoustic) on the other. The central issues of both phonology and phonetics involve deciding exactly what kinds of basic units and organizing principles exist in these three domains and how they relate to one another. The following discussion presupposes a “default” phonological organization of an ut-

terance as a distinctive feature matrix (Chomsky and Halle, 1968). Given this representation, the central problem of phonetics concerns the way in which physical properties relate to distinctive features and phonetic segments.¹

A. Strong theories

1. Double-strong theory

One important approach to this problem can be referred to as a *double-strong* theory of phonetic specification. Here the term *strong* is taken to imply a simple, robust, and transparent relation between physical and symbolic elements. Stevens and Blumstein's (1981) model is *double-strong*, in that it postulates (i) strong relations between symbols and gestures and (ii) strong relations between symbols and auditory properties. Note that there is a symmetry among objects in all three domains as represented in Fig. 1(a).

2. Strong gestural and strong auditory theories

Other theories are not so symmetrical: They postulate strong relations between *only one* of the physical domains and the symbolic level. *Strong-gestural* approaches may be exemplified by the motor theory of Liberman and Mattingly (1985). This postulates a strong, transparent relationship between symbols and gestures. There is a more complex and unidirectional path that relates gestures to auditory properties, as represented by the broken-shafted, single-headed arrow in Fig. 1(b). This reflects the complex, typically nonlinear mapping that psychomotor and biomechanical factors impose between natural units of the gestural domain and their acoustic consequences. The process of the inverse mapping from acoustics to symbols is not depicted. According to Liberman and Mattingly, this likely involves a complex analysis-by-synthesis process that generates internal auditory matches from symbolic hypotheses using an internal model

^{a)}“Selected research articles” are chosen occasionally by the Editor-in-Chief, that are judged (a) to have a subject of wide acoustical interest, and (b) to be written for understanding by broad acoustical readership.

| | | | | | |
|----------|---|----------|------|----------|---------------------|
| Gestural | ↔ | Symbolic | ↔ | Auditory | (a) Double-strong |
| Symbolic | ↔ | Gestural | ---> | Auditory | (b) Strong gestural |
| Symbolic | ↔ | Auditory | <--- | Gestural | (c) Strong auditory |
| Gestural | ← | Symbolic | ← | Auditory | (d) Double-weak |

FIG. 1. Relations among domains in several theories. Double-shafted arrows indicate strong relations. Double-headed arrows indicate bidirectional relations. Broken-shafted arrows indicate highly complex, indirect relationships. Solid single shafted arrows indicate moderately complex, but highly systematic relations.

of the unidirectional mapping postulated. The direct realist theory of Fowler (1989) can probably also be represented fairly by Fig. 1(b). However, the inverse mapping from acoustics to gestures is *not* assumed to require analysis-by-synthesis decoding, but instead is said to involve “direct perception” of gestures by listeners.

Strong-auditory theories, exemplified most clearly by Diehl, Kingston, Kluender, and their colleagues (Diehl and Kluender, 1989; Kingston and Diehl, 1994, 1995) hold essentially the opposite position from that of the strong gesturalists. Relations are assumed to be strong between symbols and auditory properties but only weak and indirect between symbols and gestures [see Fig. 1(c)].

B. Criticism of strong theories

1. *Contra auditorists*

Any argument against either a strong auditory or a strong gesturalist approach is *a fortiori* an argument against a double-strong theory, so it suffices to review those two cases. The gesturalist criticism of auditorists has a long history. A main line of argumentation has been that the complex web of acoustic cues relevant to a particular phonetic distinction can only be understood in light of its articulatory source. An impressive collection of trading-relation and multiple cue experiments from the Haskins group have emphasized this theme (e.g., Liberman and Mattingly, 1985; Repp, 1982).

Consider the case of the voicing distinction in English stop+ vowel syllables. Properties favoring voiced stops include negative or zero voice onset time (VOT), low *F1* onset, and lowered *F0*. Those favoring voiceless stops include positive VOT, high *F1* onset, and raised *F0*. These patterns are often interpreted by gesturalists as natural byproducts of a single glottal timing gesture. [See Kingston and Diehl (1994) for a critical review.]

2. *Contra gesturalists*

Auditorist criticism of gesturalist claims has a fairly long history in the case of vowels.² Ladefoged *et al.* (1972), Nearey (1980), and Johnson *et al.* (1993) have all argued that articulatory targets for the same vowel are quite varied across speakers. Since the corresponding acoustic output is

more nearly invariant, they argue that linguistically relevant properties of vowels are acoustic or auditory rather than articulatory. Perkell *et al.* (1993) have recently shown a kind of trading relation between tongue and lip position in the production of /u/ that seems to be motivated by the acoustic synergy of distinct labial and velar constrictions. Some speakers appear to use varying articulatory means to achieve a relatively constant acoustic end in different phonetic contexts.

Kingston, Diehl, Kluender, and their colleagues (e.g., Kingston and Diehl, 1994, 1995; Diehl and Kluender, 1989) have recently launched a barrage of attacks on the gesturalist position. They have argued that a number of properties claimed to be the natural fallout of the biophysical interaction of gestures instead result from *deliberately controlled* actions. Specifically, they argue that many properties that covary in production are *actively managed* to produce auditory enhancement through the creation of a set of acoustic subproperties whose combination leads to derived perceptual properties (e.g., the *C:V ratio* or the *low-frequency property*) to which human auditory systems are particularly sensitive.

We thus have entirely contradictory claims from auditorists and gesturalists as to the organizational basis of important context dependent cues. As I have noted elsewhere (Nearey, 1991), I am genuinely impressed by the quality of the research by both the auditorists and the gesturalists that is critical of the other position. Each has convinced me that the others are wrong.

C. Simultaneous constraints in acoustic-articulatory space

Proponents of the double-strong theory are manifestly concerned with both articulation and audition. Although primarily defined in auditory terms, even the original Jakobsonian features (Jakobson *et al.*, 1963) were also supplied with straightforward (though sometimes not gesturally unique, as in the case of [±flat]) articulatory implementations. The development of the successor to this theory by Stevens and his colleagues has been directed toward finding a kind of harmony between stable articulatory implementations and robust auditory properties (Blumstein and Stevens, 1980; Stevens and Blumstein, 1981; Stevens and Keyser, 1989; Stevens, 1990). Stevens’ quantal theory suggests the motivation for the selection of a universal set of binary features rests on what might be considered “sweet spots” (Nearey, 1995) in the gestural by auditory space.

While simultaneous concerns for articulation and acoustics are natural to the double-strong approach, researchers in the other strong camps have also made concessions in that regard. Thus, the auditorists Kingston and Diehl (1994) argue that speakers “...optimize their phonetic behavior by both minimizing articulatory effort and maximizing [auditory] distinctiveness (p. 423).” Conversely, the gesturalist Fowler (1989) has stipulated that “perceptual constraints guide the development of sound inventories and of phonological processes in languages (1989:145).”

Authors from both the single-strong schools have on occasion approvingly cited the research of Lindblom (e.g., 1986, 1990). Lindblom’s work partly involves the issue of

real-time phenomena that affect immediate communication (e.g., his hyperspeech and hypospeech).³ However, it also involves longer-term considerations such as the acquisition of sound contrasts and the truly secular issue of (sociocultural) evolution of phonological inventories. The distinction between long-term (diachronic), acquisitional, and real-time constraints may be crucial for understanding many aspects of linguistic behavior, including speech perception.

D. A double-weak approach to symbol-signal mapping

The *double-weak* approach outlined below can be viewed as a fallback theoretical position. In principle, the other theories (all of them, but double-strong theory in particular) would be preferable if they could be reconciled with the data. However, until there is compelling evidence that this can be done, it seems advisable to entertain other possibilities.

Figure 1(d) depicts the relation among domains in double-weak theory. Only two conditions are necessary for speech to operate as an effective communication system. First, a symbol sequence must be encoded into gestures. Second, the acoustic output of those gestures must provide the listener with auditory cues sufficient to decode the intended symbol sequence. Only subsets of symbol-to-gesture and sound-to-symbol mappings that meet this condition can be considered. On this account, gestures and auditory properties are linked only indirectly, through separate links to shared symbols. The arrows are unidirectional and are drawn with thin shafts in Fig. 1(d) to indicate that, unlike the strong theories, the relationships of properties to symbols is not necessarily straightforward. They are assumed only to be tractably systematic, to lie within the range of the feasible for the (possibly speech-specialized) auditory and motor control systems. Longer-term and secular trends effectively impose a communicative natural selection, ensuring that the phonology of a language remains within easy reach of the vast majority of speakers and listeners.

1. Double-weak phonetics from double-strong antecedents

It is instructive to imagine how a double-weak phonetic system might emerge from a double-strong one. Consider the case of the phonological opposition between voiced and voiceless stops postvocally in prepausal position, i.e., in sequences such as /Vt/ or /Vd/. The textbook articulatory implementation of this opposition involves the presence (versus absence) of phonation during closure. This is a *double-strong* scenario in that both the gestures and their key acoustic consequences are readily identifiable. The portion of the waveform associated with the open vocal tract is readily distinguishable from the quasisinusoidal voice bar of the closed vocal tract section. As the vocal tract constriction becomes more radical and approaches complete closure, *F1* will approach its low-frequency, closed tract asymptote (Fant, 1960).

Consider a set of hypothetical diachronic changes that could lead to a cue pattern more like that of modern English. For well-known aerodynamic reasons (Ohala, 1981; Kingston and Diehl, 1994), it is relatively difficult to produce

voicing during complete vocal tract closure and some nearly heroic maneuvers may be necessary to maintain it for long. On the other hand, for the voiceless stops, glottal pulsing may not stop immediately when the vocal tract closes without additional adjustments (Kingston and Diehl, 1994). To maintain sufficient distinctiveness (Lindblom, 1990; Martinet, 1955; Kohler, 1984) some speakers may shut down voicing actively, e.g., by abducting the vocal folds a little early (or fully abducting them to form a glottal stop). This will keep voice bar out of the closure period. It will also cause a change in signal type at the time just prior the vocal tract closure. Strictly speaking, the result would be preaspiration (or glottalization). It may, however, be very close in its acoustic effect to silence. A necessary side effect of this gesture will be the truncation of the falling *F1* pattern, as oral closure now occurs after *F1* has become inaudible.

On the basis of results of Watson and his colleagues (Kewley-Port *et al.*, 1988; Watson and Kewley-Port, 1988; Watson and Foyle, 1985), it seems reasonable to suggest that the human auditory system is fully capable of tracking most of these modifications in good listening conditions. Assume also that listeners are capable of exploiting the correlation between apparent vowel duration, apparent closure duration, *F1* termination, and the voiceless character of the following closure interval to support the phonological contrast between the final stops.

From a strong-gesturalist perspective, all the consequences of the change in glottal timing would be attributed by listeners to just that. Under a double-weak account, however, the cues have a potentially independent status and may be seized upon by other speakers *as proper signals of the consonantal opposition*. This, in turn, opens the door for some of these cues to be approximated *by means quite distinct from glottal timing*. Thus, the most salient effects of preaspiration (or glottalization) might be approximated by a longer (unvoiced) oral closure period. Such programmed changes in gestural patterns might further shift the “functional load” of cues so that the presence of voice bar during closure is not as important as it was earlier and speakers may be able to get away with hardly producing any at all.

Other noticeable properties that may have started as accidents of production could also be enhanced. Vowel shortening is a likely articulatorily byproduct of early devoicing and of closure enhancement in the voiceless environment. However, the duration of the voiced closure could also be shortened for /Vd/ (and the V actively lengthened) to enhance the emerging differences in temporal signature between the /Vt/ and /Vd/ patterns.

Although some speech-specific perceptual mechanisms might be involved, it seems more reasonable to assume that such a reweighting of cues could take place through a general process of auditory-perceptual learning. Cue evolution is viewed as diachronic in the traditional sense. As such, children and perhaps older speakers who are retuning their dialect under sociolinguistic influences (Labov, 1972) might be expected to bear special responsibility in these changes. This resembles Ohala's account of listener-oriented sound change (Ohala and Shriberg, 1990). However, unlike the cases considered by Ohala, all the changes just discussed are strictly

subphonemic, operating in effect only on cue weights. There is no quantal reinterpretation at the symbolic level, only the specific gestural and acoustic properties that map the symbols have changed. At some later stage, if voice bar were eliminated during /d/-closure and if (perhaps later still) *F1* termination differences were eliminated, phonologists would likely reinterpret the differences in terms of distinctive consonantal or vocalic quantity (with the other redundantly varying), rather than as a voicing distinction. Language acquires might arrive at a similar solution. At this point, an Ohala-style phonological *hypocorrection* might be said to have occurred.

2. Double-weak versus strong auditorist accounts

This account differs from that of strong-auditorists primarily in that it concedes that a principle source of the kinds of covariation in cue patterns is articulatory behavior. The complex pattern of *F1*, closure duration, and vowel duration can plausibly be attributed to an initial change in glottal timing. While auditorists like Kingston and Diehl have not entirely ruled out articulatory influences on the choice of covarying patterns, they have suggested that the primary source of covarying cue patterns are what they have termed *intermediate perceptual properties* (Kingston and Diehl, 1995). Two such properties plausibly involved in the case just discussed are the *V:C ratio* (which integrates information about vocalic and consonantal duration) and the *low-frequency property* (which integrates information about *F1* offset and voice bar during closure). Such derived properties are presumed to be “hard-wired” in human (and possibly other) auditory systems, but unlike the quantal features of Stevens, they are not themselves part of a universal speech-specific inventory. Rather, Kingston and Diehl (1994) appear to assume that the derived auditory properties can be combined in various language-specific mixtures into a universal set of distinctive features.

As in Kingston and Diehl’s account, a double-weak approach assumes that temporally distributed relational features (such as the V:C ratio which can span a few hundreds of milliseconds) can be incorporated into localized phonological oppositions (such as the voicing distinction in stops). It also admits of language-specific weights. However, these weights are assumed to be applied to more primitive perceptual cues, i.e., to relatively direct perceptual correlates of acoustic properties like vocoid duration, voice-bar duration, etc. In this regard, it resembles the model of Lindau and Ladefoged (1986).

E. Pattern recognition of stylized output

A double-weak account may admit even weaker and more complex relations between phonological elements and physical properties than those described in Lindau and Ladefoged (1986), since it also allows considerable flexibility in the time alignment of properties and sharing of information among segmental elements. The theory as sketched above seems *a priori* to have little chance of offering any insight into universals of human language. [See Blumstein (1986).] From this perspective, it is precariously close to a neostruc-

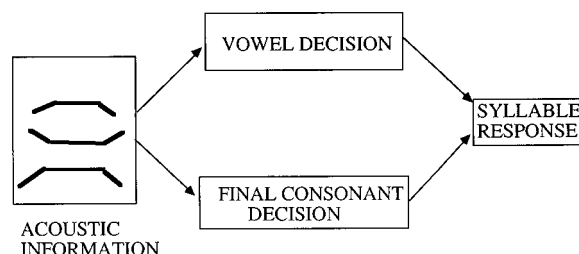


FIG. 2. Mermelstein’s 1978 model for VC judgments. Adapted from Mermelstein (1978), Fig. 3, with permission from Psychonomic Society Press.

turalist position which posits that languages can have phonologies that can vary arbitrarily from each other. Although the work of Ladefoged (e.g., 1990) and others offers strong evidence that phonologies may differ from each other in ways not contemplated by some distinctive feature philosophies, the variation observed does *not* appear to be unbounded. Without additional constraints, we might expect the development of wildly varying phonologies across languages. However, additional real-time demands of production and perception may limit the relation of symbol to signal more stringently.

What I suggest is that phonologies must allow for coarticulatory influence in production that eases the real-time demands on the speaker to move articulators too far too fast. On the other hand, coarticulatory influences may be actively managed to accommodate perceptual mechanisms of rather limited and specific computational capacity. This suggests that coarticulation is *stylized* to produce stylized covariance patterns in speech output. While the *direction* of the covariance has articulatory motivation, the details of the final patterns are constrained by what Sussman *et al.* (1995) have recently called “orderly output conditions.” Such stylized output patterns are amenable to decoding by relatively simple pattern-recognition techniques. If so, their simplicity should be revealed in the analysis of appropriately designed speech perception experiments.

II. FACTORING PHONOLOGICAL OPPOSITIONS

A. Cue sharing in VC syllables

Mermelstein (1978) presented an analysis of experimental results that lead to surprising conclusions. Mermelstein’s experiment involved VC syllables, where the vowel ranged over /ε/ and /æ/ and the consonant over /t/ and /d/. The stimuli involved manipulation of *F1* of the steady-state vocoid and vocoid duration. Mermelstein’s analysis suggested that both vowel decisions and consonant are influenced by both *F1* and vocoid duration, but *every aspect of the vowel decision is independent of consonant decision and vice versa*. (See also Allen, 1994.) This is represented schematically in Fig. 2.

The kinds of questions Mermelstein raised in his experiment can be addressed directly using logistic regression analysis, which shares many characteristics with the analysis of covariance (Haberman, 1979; McCullagh and Nelder, 1989). Two general classes of effects can be isolated in these models, as shown in Table I. These are bias effects and

TABLE I. Effects in a logistic regression model for Mermelstein's 1978 experiment.

| |
|---|
| Bias effects (Response only effects. Stimulus-independent effects) |
| V=vowel bias main effects |
| C=consonant bias main effects |
| V×C=diphone bias effects |
| Stimulus-tuned effects (Response-by-stimulus interactions) |
| F1-tuned effects: |
| V×F1=F1-tuned vowel effects (ε/æ distinctions) |
| C×F1=F1-tuned consonant effects (t/d) |
| V×C×F1=F1-tuned diphone effects |
| Vocoid Duration-tuned effects: |
| V×vocoid duration=duration-tuned vowel effects |
| C×Vocoid duration=vocoid duration-tuned vowel effects |
| V×C×vocoid duration=vocoid duration-tuned diphone effects |

stimulus-tuned effects. The latter may be further broken down for each stimulus dimension. (For a more detailed discussion, see Nearey, 1990.)

A very simple (simpler than Mermelstein's) model that allows no property-sharing can be contemplated. Here, the F1 variation is related to vowel choices only, while the consonant decision is related to vocoid duration only. This can be called a *primary cue model*.⁴ It would contain only the following terms: V, V×F1, C, C×vocoid duration. A labeled partitioning of the pattern space consistent with such a model is shown in Fig. 3(A). Such *territorial maps* can be generated in ways detailed by Nearey (1990, 1992) from the coefficients of the logistic regression models. Note that the vowel category boundary depends on (or "is tuned by") F1 but is independent of vocoid duration. Similarly, the consonant boundary is tuned by duration, but is orthogonal to F1.

Mermelstein's model in Fig. 2 could be termed a *secondary cue model*, where duration serves as a secondary cue to the vowel distinction and F1 serves as a secondary cue to the consonant distinction. This model is characterized by a logistic model including the following effects: V, V×F1, V×vocoid duration; C, C×F1, C×vocoid duration. The territorial map for such a model is shown in Fig. 3(B). Notice here that there is only a single line separating /Vt/ and /Vd/ responses and a single line separating /εC/ and /æC/ responses. Since neither of the lines are orthogonal to either of the axes, both consonants and vowels can be said to be *tuned* by both stimulus properties.

Whalen (1989) reported a replication and extension of Mermelstein's experiment and provided an analysis that indicated that Mermelstein's results did not generalize to larger scale experiments. Nearey (1990) reanalyzed Whalen's data using logistic regression analysis and agreed with Whalen that Mermelstein's model was not adequate. However, Nearey's reanalysis showed that Mermelstein's model required only slight modification to achieve very good agreement with the data. This modification was the addition of *diphone bias* effects (V×C). A key property of such models is that syllables differing by a single phoneme are constrained to have *parallel boundaries*: The /εt-æt/ boundary must be parallel to the /εδ-æd/ line, and the /εt-εδ/ boundary must be parallel to the /æt-æd/. The magnitude of the diphone bias effects required by Whalen's data are actually

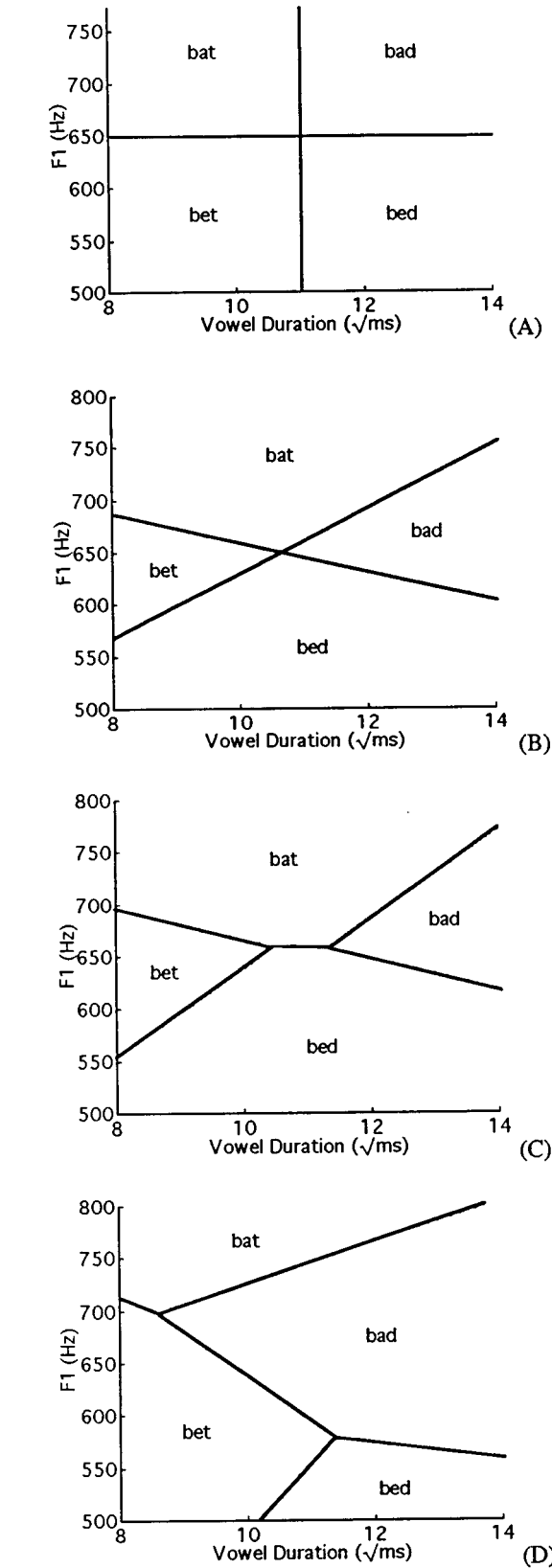


FIG. 3. Territorial maps for primary cue model (A); Mermelstein-like secondary cue model (B); secondary cue model with diphone bias terms (C); and full diphone model including stimulus-tuned diphone terms (D).

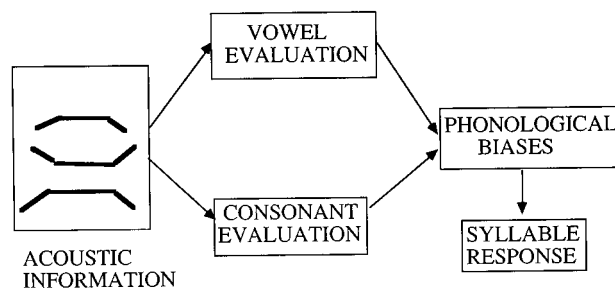


FIG. 4. Modification of Mermelstein's model to accommodate Nearey's (1990) diphone biases.

quite small. Their main consequences can be seen by comparing Fig. 3(C), which depicts the territorial map implied by Nearey's selected model, to the simpler secondary cue model of Fig. 3(B). Conceptually, the necessary modification to Mermelstein's block diagram is illustrated in Fig. 4. Note that consonant evaluation and vowel evaluation are modified by phonological (e.g., diphone) biases before a syllable decision is made.

What might diphone bias effects represent? They may serve several logically distinct functions. First, they can play the same role in implementing phonotactic constraints as Massaro and Cohen's (1983) contextual features do in their fuzzy logical models (which are computationally very similar to the logistics discussed here). Second, they may "absorb" differential lexical bias effects, whereby certain words (or nonsense syllables) are preferred by listeners independent of stimulus properties. Finally, they may serve simply as *fudge factors* that make the model fit the data better. In this example, the net effect of the diphone bias term is to increase the area of "bat" and "bed" responses (symmetrically, as it turns out for technical reasons) at the expense of the /ɛt/ and /æd/ categories.

Why might this be helpful? Peterson and Lehiste (1960) report that duration is about 40% greater for /æ/ than for /ɛ/. Similarly, vowels before /d/ are between 40% and 50% longer than those before /t/. Thus, the duration difference associated with the consonants is about the same as that associated with the vowels so that the duration pattern of the four syllables is /ɛt/ < /ɛd/ ≈ /æt/ < /æd/. In the extreme cases, the duration characteristics associated with the vowels and consonants are synergistic, both short for /ɛt/ and both long for /æd/. The middle duration region represents a kind of conflict between duration patterns of the consonants and vowels involved. An optimal decision rule should reflect some kind of correlation between consonant and vowel decisions in the middle range, where only /ɛd/ and /æt/ are likely. For example, assume that other cues were relatively neutral with respect to /-t/ and /-d/ and that the vocoid duration was in the ambiguous range. Then, if the vowel sounds like /ɛ/, a listener could expect to reduce errors by choosing /-d/, because /ɛt/ is very unlikely in this duration range. There are a number of possible ways in which such interdependence could be implemented, as discussed by Nearey (1990, 1992). One of the simplest is through adjustment ("fudging") of diphone biases to approximate the following rule: Bias decisions slightly in favor of the "conflicting-

cue" /ɛd/ and /æt/ syllables over the others. Although it may appear inelegant, such a rule may provide a simple approximate solution to an otherwise difficult problem, as it can be applied automatically and globally. (In cases where cues are not conflicting, they will simply override the mild biases in favor of the conflicting cue categories.)

It might be objected that the effect of the above analysis is not very different from simply precompiling all the stimulus differences among the four syllables into four distinct diphone patterns. Nearey (1990) argues that if such precompilation were necessary, then radically more complex patterns could be expected to occur in perception. For example, a pattern like that in Fig. 3(d) would be admissible, where there is no requirement of parallelism on the boundary lines. This corresponds to a *complete diphone* model, which includes *stimulus-tuned diphone terms* corresponding to $V \times C \times \text{vocoid duration}$ and $V \times C \times F1$ interactions. Each cue is in effect allowed to have an arbitrary weight for each diphone element. Nearey (1990, 1992) has argued that the *absence of such complex patterns* is significant and may reflect the operation of relatively limited perceptual mechanisms that structure phonological systems. However, it is possible that the parallelism observed in this example is merely accidental. Perhaps if we looked at a larger problem, more complex patterns, such as that illustrated in Fig. 3(D), might yet arise. The following, larger experiment may provide more suitable testing ground for evaluating models.

III. EXPERIMENT I: PERCEPTION OF A LARGE /hVC/ CONTINUUM

A. Method

1. Stimuli

A set of 480 /hVC/ stimuli was synthesized, where V ranged over the five Canadian English back and central vowels /u, o, ʊ, ɒ, ʌ/ and C range over /t/ and /d/. The stimuli were synthesized at 10 kHz on an implementation of the Klatt80 synthesizer (Jamieson *et al.*, 1989; Klatt, 1980). They began with a 60-ms /h/-like fricative with formant frequencies of the following vocoid, but excited only by the aspiration source. This was followed by a variable steady-state vocoid with $F3$ fixed at 2350 Hz and $F4$ and $F5$ fixed at 3300 and 3750 Hz, respectively. The vocoid had a falling $F0$ contour from 125 to 100 Hz over the course of its duration. This was followed by 52-ms transition in $F2$ and $F3$ appropriate for a coronal stop. The $F2$ values moved from the steady state of the vocoid to a preclosure target of 1040 Hz + 50% of the steady-state value. $F3$ moved from the steady state $F3$ of the vowel to 2700 Hz. $F1$ remained at its steady-state frequency during the $F2$ – $F3$ transitions.⁵ A weak, nonvoiced coronal stop burst was synthesized at 90 ms after the offset of the vocoid, with any interval between the voicing offset (of the vocoid or of the voice bar) being filled with silence.

The variable stimulus factors were arrayed in a four-factor, fully crossed design. (1) Vocoid duration controlled the duration of the steady-state vocalic part of the signal. This duration ranged from 90 to 218 ms in 32-ms steps. (2) Voice bar duration controlled a period of quasisinusoidal voicing (using the Klatt AVS parameter and $F0$ of 100 Hz).

This followed the vocoid and ranged from 0 to 90 ms in 30-ms steps. (3) *F1* controlled the first formant frequency of the vocoid and ranged from 330 to 730 Hz in eight steps. (4) *F2* separation controlled the difference *F2* minus *F1* of the vocoid and ranged from 425 to 700 Hz in three steps.

2. Subjects

Fifteen paid subjects, all of whom had a small amount of phonetic training, were recruited as listeners. Each categorized each stimulus five times, for a total of 2400 responses per subject. Responses were gathered for each subject over several sittings, usually on different days.

3. Presentation

The stimuli were presented under computer control on a PC using a 12 bit D-A converter and an antialiasing filter with a 5-kHz cutoff. Signals were then amplified and played through a small loudspeaker mounted on a table in front of the listener. The listener was seated in a sound-attenuated booth and responded using a computer mouse. The PC screen was positioned outside the booth and was visible through a double-gazed window. There were 10 response boxes labeled with the words “who’d,” “hood,” “hoed,” “hawed,” and the pseudoword “hud.” Pseudophonetic transcriptions were also provided using ASCII characters and a sheet listing the IPA transcription of the words was available for consulting in the listening booth. Subjects were run one at a time, with separate randomizations for each subject.

B. Results and discussion

1. Background

It is reasonable to divide the cues into primary and secondary for consonants and vowels, based on prior findings, including those of Nearey (1990). Primary cues are those that are expected to have a large effect on the probability of responses to the category in question. Thus *F1* and *F2* separation are both expected to strongly affect vowel responses (lower *F1* favoring higher vowels, lower *F2* backer or rounder vowels). Similarly, voice bar duration and vocoid duration are expected to strongly affect consonant response probabilities (longer voice bar and longer vocoid duration favoring /d/).

On the basis of Mermelstein’s and Whalen’s experiments (among others), vocoid duration is expected to have an effect on vowel responses (though probably somewhat weaker than *F1* or *F2*) and is classed as an anticipated *secondary* vowel cue. Similarly *F1* [both at the termination of the vocalic section and earlier in the steady-state vocoid (Summers, 1987, 1988)] is also expected to affect consonant choice (lower *F1* favoring /d/). Two other cue relations of the same formal complexity (stimulus-by-segment interaction in analysis of covariance notation) as the primary cues are included for completeness. These will be referred to as *minor cues*. They are voice bar as a vowel cue and *F2* as a consonant cue. A summary of these effects and their abbreviated labels are presented in Table II. Bias terms (whose labels involve V and C terms only, and no stimulus effects) are also indicated there.

TABLE II. Breakdown of effects in logistic models for experiment I. All effects listed are allowed by Nearey’s diphone-biased segmental models. If the VC diphone bias term is not included, models are compatible with Mermelstein’s independent segment models.

| Biases: |
|---|
| Segmental V=vowel; C=consonant |
| Diphone: VC=vowel×consonant |
| Anticipated primary cue effects: |
| Primary vowel effects: VF=V× <i>F1</i> , VS=V× <i>F2</i> separation |
| Primary Consonant effects: CB=C×voice bar, CD=C×vocoid duration |
| Anticipated secondary cue effects: |
| Secondary vowel effects: VD=V×vocoid duration |
| Secondary consonant effects: CF=C× <i>F1</i> |
| Possible minor cue effects: |
| Vowel effects: VB=V×voice bar duration |
| Consonant effects: CS=C× <i>F2</i> separation |

A model consistent with that proposed by Mermelstein (1978) could contain all the effects in Table II, except for the diphone bias terms, VC. The presence of the latter effects implies that vowel and consonant decisions are *not fully independent*. Nearey (1990) allows VC effects in what are termed diphone-biased segmental models. For the present experiment, inclusion of these terms results in a modest complication of the model requiring only an additional four degrees of freedom beyond a pure secondary-cue Mermelstein-like model.

There are, however, a number of effects (not shown in Table II) that Nearey’s (1990) hypothesis specifically disallows. These are stimulus-tuned diphone terms, represented by stimulus factors multiplying V×C diphone interactions. Thus a term like VCF would represent a vowel×consonant ×*F1* interaction.

Some proposals in the literature imply just such interactions. Fisher and Ohde (1990) and Summers (1988) suggest that the effectiveness of *F1* (both at its termination and during steady state) as a cue to a voicing may be influenced by the overall *F1* level of the preceding vowel. Thus, a small change in *F1* frequency might have a larger effect on the relative probabilities of /t/–/d/ choices in /hɒ-/ contexts where the average *F1* level is higher than in /hu-/ contexts where it is lower. Such cue-weight differences would lead to non-null V×C×*F1* effects in the modeling scheme investigated here. Similarly, arguments by Kingston and Diehl (1994, 1995) imply the existence of noticeable V×C×voice bar duration interactions. These issues will be pursued in the analysis below.

2. Logistic regression analysis

Table III focuses on the relative goodness of fit of several models. It follows Nearey’s (1990) analysis of Whalen’s “bad–bet” experiment.⁶ Column I gives a descriptive label and column II shows the effects included in the model or those added to the model in the previous row. Thus, model 3 includes all the terms in model 2 (V, C, VF, VS, CB, CD) plus the V×vocoid duration and C×*F1* terms (VD, CF). Column III reports the deviance statistic, *G*². This is the “lack of fit” measure (the smaller the value, the better the fit) typically used in logistic and log linear modeling (McCullagh and Nelder, 1989). Column V reports rms error

TABLE III. Goodness-of-fit measures of selected logistic models for experiment I.

| I Model label | II Effects included | III Deviance (G^2) | IV Residual degrees of freedom | V rms error | VI Reduction of residual deviance |
|------------------------------------|---------------------------|------------------------------|--------------------------------------|-------------------|--|
| 1. Bias only model | V C | 98 284.18 | 4315 | 21.96% | ... |
| 2. Primary cues segmental | V C VF VS CB CD | 16 099.39 | 4305 | 9.73% | 83.6% |
| 3. Secondary cue | 2+ VD CF | 5943.07 | 4300 | 4.67% | 63.1% |
| 4. Diphone biased secondary cue | 3+ VC | 5687.45 | 4296 | 4.53% | 4.3% |
| 5. Diphone biased full cue | 4+ VB CS | 5659.82 | 4291 | 4.51% | 0.5% |
| 6. Full diphone | 5+ VCF VCS VCD VCB | 5273.08 | 4275 | 4.18% | 6.8% |

(in percentage points of response probability) of predicted versus observed values. Column VI reports the percentage of reduction residual deviance (column III) from the previous line of the table. Thus the change in deviance from the primary cue model of row 2 to the secondary cue model of row 3 represents a change of $(16\,099.39 - 5943.07)/16\,099.39 = 63.1\%$.

Model 1 serves to provide a baseline error measure. It is a random guessing model with no stimulus effects. It predicts a fixed probability of choosing each vowel and each consonant category regardless of the stimulus. Model 2 is the primary cue model, where consonants and vowels have non-overlapping cues: $F1$ and $F2$ separation tune vowel decisions and voice bar duration and vocoid duration tune consonant decisions. The rms error rate has decreased by more than 12 percentage points and the residual deviance is reduced by more than 83%.

The move to the secondary cue model (model 3) reduces the absolute rms error to less than 5%. Here, in addition to the primary cues of the previous models, $F1$ is allowed to tune consonant decisions (CF) as well as vowels (VF) and vowel duration is allowed to tune vowel decisions (VD) as well as consonants (CD). This represents a purely separable Mermelstein-like model where consonant and vowel decisions are entirely independent, though $F1$ and vowel duration cues are shared. The addition of the secondary cue terms reduces the rms error by another five percentage points and reduces the residual deviance from model 2 by more than 63%. There is relatively little error left to explain, with only about 6% of the residual deviance from model 1 remaining.

The addition of the Nearey (1990) diphone bias terms VC in model 4 reduces the rms error by less than two tenths of a percentage point. The VC terms may represent “fudge factors” that might “boost” probabilities of categories with conflicting cues. They may also absorb lexical bias effects (some of the /hVC/ syllables are common words, some are not). The reduction in rms values is certainly not impressive. However, the model is already fitting rather well in absolute terms and the reduction of the residual deviance is a little more than 4% on only four additional degrees of freedom.

The addition of the two “minor cues” in model 5 (V \times voice bar duration and C \times $F2$ separation) improves the fit

very little, less than 1% additional deviance being accounted for with five additional degrees of freedom in the model.

Finally, the move from model 5 to the (complete diphone) model 6 results in a reduction of less than four tenths of a percentage point in the rms error. This model allows each diphone choice to find an arbitrary weight on each of the stimulus properties. Thus, it would be possible, if the patterns in the data warranted, for the model to attach large weights to $F1$ for the / ϕ t/-/ ϕ d/ distinction while attaching a zero weight to $F1$ for the / ϕ u/-/ ϕ u/ distinction. This additional flexibility contributes about a 7% reduction of the residual deviance. This seems a relatively modest gain given the 16 additional degrees of freedom used by the model.

3. Statistical hypothesis tests

Two types of statistical analysis are described below. The first involves quaslikelihood procedure described by McCullagh and Nelder (1989) and used by Nearey (1990). This focuses on the change in goodness-of-fit of successive models in Table III.⁷ According to this procedure, the improvement from model 1 to 2 is highly significant [$F(10,4305) = 6265.9$, $p < 0.000\,01$], as is that from model 2 to 3 [$F(5,4300) = 212.36$, $p < 0.000\,01$] and that from model 3 to 4 [$F(4,4296) = 6.475$, $p < 0.000\,04$]. The addition of the minor cue terms in model 5 is not significant [$F(5,4291) = 0.552$, $p < 0.5$]. However, the addition of the stimulus-tuned diphone terms in model 6 is significant [$F(16,4275) = 2.281$, $p < 0.003$].

The second set of tests assesses the reliability of logistic coefficients across subjects. The most complex model (model 6) is first fit to the data of each subject individually. The coefficients from these analyses are then subjected to the second-stage multivariate tests described in Gumpertz and Pantula (1989).⁸ The results of these tests are summarized in Table IV.

The general pattern of results agrees with the quaslikelihood analysis. The primary cue (VF, VS, CB, CD) and secondary cue (CF, VD) effects are all highly significant. The minor cue effects (VB and CS) are not significant. The diphone bias effects (VC) are significant. Three of the four diphone-tuned stimulus effects (VCF, VCD, and VCB) are

TABLE IV. Random effects *F* tests for families in model 6 of Table III.

| | Family | <i>F</i> | <i>df</i> | <i>p</i> |
|----|--------|----------|-----------|----------|
| 1 | V | 87.199 | 4, 11 | 0.000 00 |
| 2 | C | 58.227 | 1, 14 | 0.000 00 |
| 3 | VC | 6.783 | 4, 11 | 0.005 27 |
| 4 | VF | 102.639 | 4, 11 | 0.000 00 |
| 5 | CF | 91.704 | 1, 14 | 0.000 00 |
| 6 | VCF | 6.899 | 4, 11 | 0.004 95 |
| 7 | VS | 79.881 | 4, 11 | 0.000 00 |
| 8 | CS | 1.434 | 1, 14 | 0.251 04 |
| 9 | VCS | 1.468 | 4, 11 | 0.277 23 |
| 10 | VD | 48.427 | 4, 11 | 0.000 00 |
| 11 | CD | 76.086 | 1, 14 | 0.000 00 |
| 12 | VCD | 4.261 | 4, 11 | 0.025 25 |
| 13 | VB | 1.705 | 4, 11 | 0.218 72 |
| 14 | CB | 106.104 | 1, 14 | 0.000 00 |
| 15 | VCB | 8.646 | 4, 11 | 0.002 08 |

significant, while a fourth (VCS) is not.⁹ The presence of reliable diphone bias effects is not surprising (Nearey, 1990) and has in fact been “reconciled” within double-weak theory (Nearey, 1992). The significance of diphone-tuned stimulus effects is another matter and deserves further attention.

4. Graphic analysis of logistic coefficients

Considerable insight into the nature of stimulus-response relations can be obtained by a study of coefficients of a logistic model using diphones as the basic symbolic unit. The coefficients from this model can be displayed in a manner similar to interaction plots in analysis of variance (ANOVA), as shown in Figs. 5–9. The *y* axis shows the value of the coefficients. The *x* axis represents the vowel categories roughly in order of their IPA vowel height. The two lines represent /-Vt/ and /-Vd/ syllables. Note that if consonant and vowel choices were totally independent as in Mermelstein’s original proposal, *all* of these plots would show only parallel lines. Nearey’s (1990) diphone biased models require parallel lines in all panels except Fig. 5, which displays bias coefficients.¹⁰

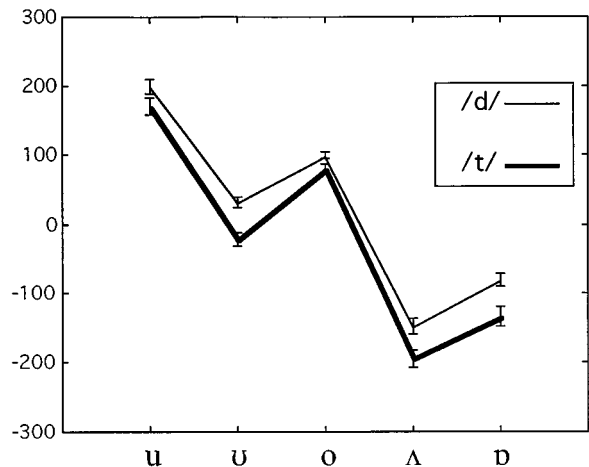


FIG. 5. Mean bias coefficients (and standard errors across listeners) corresponding to model 6 of Table III.

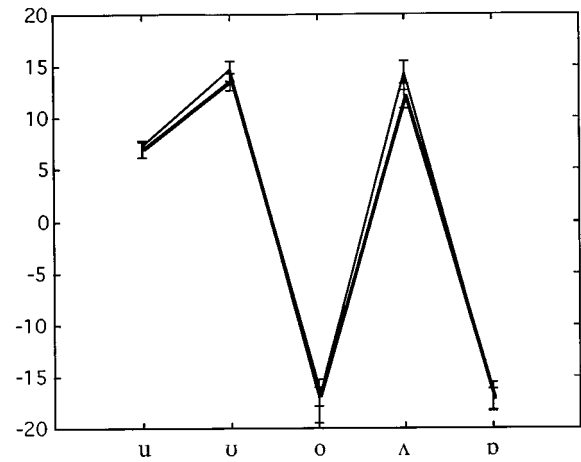


FIG. 6. Mean *F*₂ separation coefficients (and standard errors across listeners) corresponding to model 6 of Table III.

For Figs. 6–9, the ordering of the coefficient values roughly reflects the expected ordering of the categories along the stimulus dimension in question. Thus, in Fig. 8 the coefficients associated with the vowel categories increase roughly in order of expected *F*₁ associated with the categories. Similarly, the coefficients for /-Vt/ categories and /-Vd/ are ordered in the expected way (higher values associated with /-Vt/).

Figure 6 shows that the *F*₂-separation value has essentially no effect on consonant judgments, while the vowels differ substantially. This agrees with the statistical results in rows 7–9 of Table IV, where only VS effects are significant, while VC and VCS are not. The pattern of coefficients indicates that vowels /o/ and /b/ are favored by small *F*₂–*F*₁ differences and the others by larger separations. These patterns are roughly in line with production data patterns (Nearey and Assmann, 1986).

Figure 7 shows that vowel duration coefficients are noticeably different for both consonants and vowels. This accords with the significance of the VD and CD terms in rows 10–12 of Table IV. These lines appear to be roughly parallel.

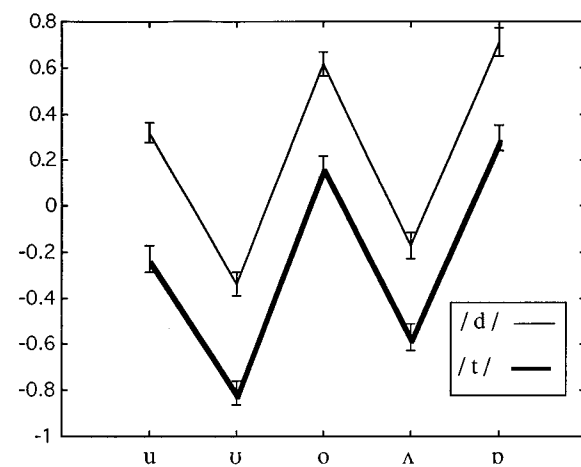


FIG. 7. Mean vocoid duration coefficients (and standard errors across listeners) corresponding to model 6 of Table III.

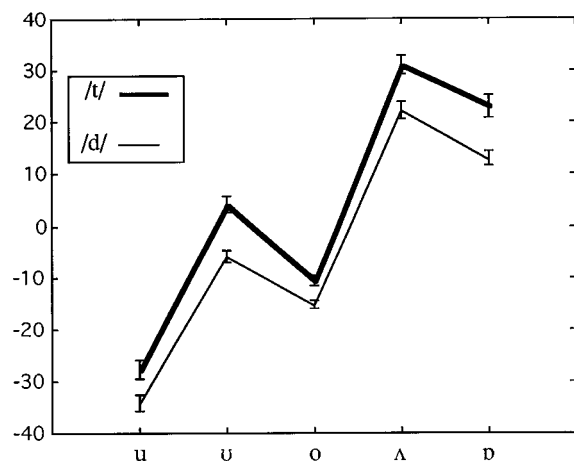


FIG. 8. Mean $F1$ coefficients (and standard errors across listeners) corresponding to model 6 of Table III.

This has the same interpretation as in similar ANOVA means plots: There are stimulus-tuned main effects for vowels and for consonants (corresponding to VD and CD terms in the models discussed above) but little evidence of important VC interactions. Nevertheless, the VCD term is nominally significant in Table IV. There appears to be a slight trend for wider separations of the /-t/ and /-d/ lines for vowels associated with lower $F1$'s. Since the deviations from parallelism are quite small and since there is no clear basis from alternate theories for the pattern observed, they will not be discussed further.¹¹

On the other hand, the VCF interaction in Table IV is significant and the $F1$ -tuned effects in Fig. 8 show reasonably clear visual evidence for lack of parallelism of the /-d/ and /-t/ lines. Furthermore, there is also discussion in the literature relevant to interactions. It has been suggested by Fisher and Ohde (1990) and by Summers (1988) that the effectiveness of $F1$ as a cue to voicing may be influenced by $F1$ level. Thus for high $F1$ vowels, variation in $F1$ within vowels might exert a fairly large influence on consonant judgment, compared to low $F1$ vowels. In Fig. 8, the effectiveness of $F1$ in separating the consonants for different

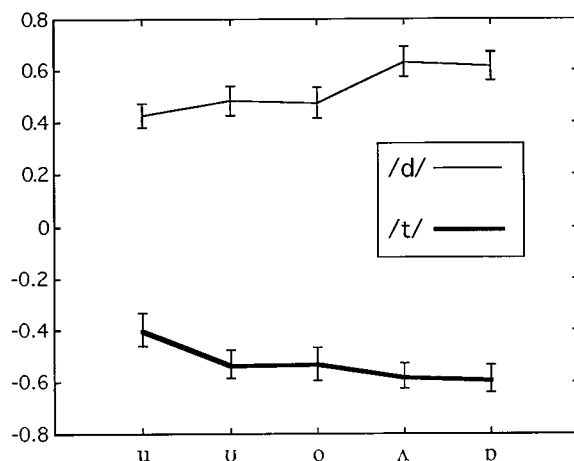


FIG. 9. Mean voice bar duration coefficients (and standard errors across listeners) corresponding to model 6 of Table III.

vowel choices is reflected by the separation of the corresponding /-Vt/ and /-Vd/ points. If attention is limited to only the lowest and highest $F1$ vowels (/u/ and /ɒ/), one might conclude that there is some support for Fisher and Ohde's hypothesis. However, the overall pattern indicates that $F1$ is relatively less effective for the vowels /u/ and /o/, while /u/ [which has comparable $F1$ characteristics to /o/ in this dialect; see Nearey and Assmann (1986)] shows a larger difference, roughly comparable to the two highest $F1$ vowels. It is possible that some more complex interaction is taking place, e.g., that when both $F1$ and $F2$ are relatively low, $F1$ has a smaller effect. However, I know of no theoretical basis for such a finding. Evidence for lack of parallelism is still not particularly strong ($F1$ noticeably affects both vowel and consonant choice throughout) and it would seem difficult to argue that the relatively subtle deviations have great theoretical import without clearer support from other experiments.

Finally, consider Fig. 9, which displays the effect of voice bar. A potential source of diphone interaction for this case can be derived from an analysis of VCV's by Kingston and Diehl (1995). They suggest that voice bar affects voicing judgments by contributing to the low frequency property in all $F1$ contexts. However, they also argue that the [\pm voice] feature is implemented in English by both the low-frequency property and by V:C ratio. Consequently, the relation of this property to voicing judgments should be nonlinearly affected by $F1$ offset. When $F1$ offsets are low and spectrally contiguous with voice bar, the perceived V:C ratio is increased because the apparent gap duration is shorter (and the apparent vowel duration longer). This implies that consonant decisions should be more strongly affected by voice bar for low- $F1$ vowels like /u/ (where voice bar contributes to both the low-frequency property and to V:C ratio) than for the high $F1$ vowels like /ɒ/. Such differential effectiveness of voice bar duration could be expected show up in Fig. 9 as a wider separation of the /-Vt/ and /-Vd/ lines for the low $F1$ vowels and narrower separation for the higher $F1$ vowels. In fact, a slight trend in the opposite direction is found. Again, I know of no theory that would predict such a pattern.¹²

In sum, it is important to note that there are some statistical and graphic indications that Nearey's (1990) diphone-biased segmental models are not entirely adequate to the data from this experiment. Nevertheless, the overall goodness-of-fit of a straightforward extension of Mermelstein's (1978) model is remarkably good (model 3 in Table III, with about 5% rms error). The degree of divergence from parallelism in Figs. 6–9 is modest and the direction of deviations noted does not appear to relate very well to any previously articulated theory.¹³

IV. GENERAL DISCUSSION

A. Pattern-recognition models for more complex data

Every language learner is faced with a *fait accompli* of what Kuhl (1992) has aptly called "the ambient language." A key task for the language learner is the construction of an appropriate pattern-recognition scheme for the problem at hand. Analysis of the stimulus-to-response mapping for mature listeners in laboratory studies suggests that the compu-

tational complexity of the adult recognition scheme is rather stringently constrained. If transmission of symbolic information from speaker to hearer is to approach optimality (i.e., if speakers produce patterns that can usually be recognized), similarly stringent constraints must apply to output patterns. The models discussed thus far in this paper are all linear in their stimulus effects. This implies that the decision regions for each category in territorial maps (e.g., Fig. 3) are linearly separable, i.e., delimited by straight lines.¹⁴ While many speech patterns may be compatible with this restriction, there appears to be at least one notable exception. This involves the cue patterns associated with place of articulation in stop consonants.

Nearey and Shamma (1987) published work on phenomena that are now referred to as “locus-equations” (Sussman *et al.*, 1995; Sussman *et al.*, 1993; Sussman *et al.*, 1991). We measured $F2$ and $F3$ of voiced stop+ vowel syllables as early as possible after the stop burst ($F2_i$ and $F3_i$) and at 60 ms into the vocalic portion ($F2_v$ and $F3_v$) of the syllable. Results from 10 speakers showed that formant frequencies clustered quite closely about a regression line for each consonant. However, for $F2$ the /d/ line crosses both of the others, clearly indicating a lack of linear separability. Furthermore, statistical tests showed that the patterns did not have equal covariance matrices for the three consonants and that quadratic, rather than linear, discriminant methods were called for. Using these, we obtained approximately 72% correct partition of the training data into the three classes.¹⁵ Although such recognition models involve territorial maps more complex than those required for experiment I, it is important to note that classification takes place without reference to stimulus characteristics of specific CV diphones. Rather, for each consonant a single mean vector and a single covariance matrix are estimated *without regard to the phonological identity of the vowel*.

Sussman *et al.* (1993) have studied locus equations in a number of different languages, finding reliable linear patterns within each language. They have also shown that language-specific differences associated with secondary-articulation factors, such as pharyngealization, have reliable effects on these patterns. Sussman and colleagues have further argued that the resulting linear patterns are not simply an accident of production, but rather that they likely represent a deliberate “steering” of coarticulation effects to produce well-defined covariance patterns. [See also Nearey (1992).]

The degree of statistical separability of the three stop classes from information represented in locus equations is clearly not enough to account for very high identification rates for stop+ vowel syllables by humans. However, other cues, including the shape of release spectra, are known to provide substantial information about place for stops (Stevens and Blumstein, 1981; Lahiri *et al.*, 1984; Forrest *et al.*, 1988). Nossair and Zahorian (1991) have demonstrated a speaker-independent automatic recognition system that achieves very high identification rates (93.7% correct on an independent test set) for the six English stops in 11 vowel environments on 15 speakers (males, females, and children). Nossair and Zahorian used quadratic discrimination methods

(based on separate mean vectors and separate covariance matrices for each consonant) using measures derived from the first 60 ms of stop+ vowel+ C syllables. The performance of the best algorithms they studied approached that of a panel of human listeners (96.6%) on the same syllables. (Listeners showed performance of only 89.9% when they heard only the first 50 ms of the stimuli, a situation more comparable to the computer modeling.) There was a reasonably good correspondence between error matrices for listeners and Nossair and Zahorian’s best pattern recognizer when errors were pooled over speakers and vowels. There was also some correlation of error rates for individual talkers.

B. Strong and weak theories of phonological contrast

Proponents of various strong theories might well object that the complexity of the signal representations and relatively large number of parameters required in Nossair and Zahorian’s model makes them unlikely candidates as perceptual models. Nevertheless, their results are compatible with an extended version of Sussman’s *orderly output conditions*. Their algorithm effectively establishes an upper bound on the computational complexity required for a creditable solution of this long long-standing perceptual puzzle.

The presence of such an upper bound should encourage the search for a better understanding of the patterns in terms of more primitive phonological and psychophysical elements, including classical distinctive features and Kingston and Diehl’s intermediate perceptual properties. However, we should also consider the possibility that acquisition of phonological contrasts involves extensive auditory-perceptual learning. Listeners may effectively gather statistics about the auditory signatures of phonological symbols from the ambient language. It is not beyond imagining that mastery of such patterns implies the equivalent of learning mean and covariance patterns for a few dozen symbols in a space of a few dozen perceptual parameters.

The degree of perceptual learning and the complexity of cue structures implied by such a program may seem wrong-headed to advocates of strong theories. Double-strong theorists clearly favor temporally localized, quantal, universal cues. Yet some definitions of features, such as the “low frequency property” (coined by Blumstein and Stevens), do take account of some lack of absolute synchronization. Furthermore, some cues proposed by Lahiri *et al.* (1984) to separate labials from coronals involve relatively complex relational properties of temporally distributed spectral patterns, information that seems hardly more restricted in scope than that used in the locus-equations models. Auditorists like Kingston and Diehl (1994) also acknowledge temporally distributed cue information involving overtly relational properties. However, they favor an inventory of universal derived auditory properties which are functions of multiple subproperties which stand in an “interlocking network of mutual enhancement relations” to the derived properties.

Both auditorists and double-strong theorists allow for some measure of perceptual learning (see, e.g., Stevens and Blumstein, 1981) and the differences among the alternatives might best be viewed as a matter of degree. Dealing with a limited inventory of universal distinctive features or of inter-

mediate perceptual properties would pose no problems in principle for the kind of perceptual modeling I am advocating. Indeed, if they could be established firmly, they would aid perceptual modeling by limiting the degrees of freedom of the cue weights or statistical distributions to be estimated. However, a premature commitment to a limited set of properties may jeopardize modeling accuracy and, in a rush to universals, we may miss some very important facts about how language works.

In light of this danger, it seems appropriate that at least some of us pursue what may seem to strong-theorists to be a crassly empirical approach to phonological contrast. Yet, something similar has been already been advocated by Kohler (1981) and by Ladefoged (1990; Lindau and Ladefoged, 1986). There is considerable evidence from cross-language studies of production (see, e.g., Lindau and Ladefoged, 1986; Ladefoged, 1990; Sussman *et al.*, 1993) that phonetic patterns of language differ from each other in subtle and complex ways. There is also evidence from perception that speakers of different languages differ in the weights they assign to different cues (Crowther and Mann, 1992; Munro, 1992; Kuhl *et al.*, 1992; Kohler, 1981; Flege *et al.*, 1994; Scholes, 1967). Until a stronger theory is able to present a far more compelling case than any has so far, it may be prudent to take such differences at their face value, allowing language-specific patterns to be incorporated directly into perceptual models of phonological contrast. In due course, the language-specific empirical generalizations summarized by such models may serve as grist for the mill of a stronger and more convincing theoretical approach yet to be discovered.

ACKNOWLEDGMENTS

Portions of this paper were presented at the 129th meeting of the Acoustical Society of America, June 1995 in Washington DC. This work was supported by a grant from SSHRC.

¹See Nearey (1995) and Massaro and Oden (1980) for discussion of this assumption. A number of other important assumptions are also being made but not defended. Two of the more salient are: (1) The segmentation of the input signals and measurement of stimulus properties before they are presented to a perceptual model. (2) What we learn in the laboratory generalizes to more complex speech communication situations.

²For a more general discussion of the role of articulation in speech, see McGowan and Faber (1996) and related papers in the same issue.

³Lindblom's (1990) research on hypospeech suggests that a substantial part of the variation associated with weaker prosodic contexts may represent *true information loss* at the phonetic level that can be tolerated because of higher-level redundancies. See also Allen (1994).

⁴The distinction between "primary cue" and "secondary cue" is probably just a matter of degree. Cues with higher weights are relatively primary.

⁵ F_1 offset was not varied as a separate factor because of the large number of stimuli involved. Compromise values were considered in initial planning stages, but stimuli without final F_1 transitions sound quite acceptable (although a weak thump can be heard in stimuli with high F_1 's if one listens carefully). Fortunately, these stimuli are very well suited for comparison with some of Kingston and Diehl's observations about possible interactions of voice bar with F_1 level. See Nearey (1995) and note 12.

⁶Stimulus dimensions were transformed by taking logs of frequencies and square roots of durations before analysis. Experience with data analysis for vowels indicates that log frequency works about as well as any other proposed tonotopic scales (e.g., ERB, Bark), all of which work better than

linear (Hz) measures. Square roots of durations were also used by Nearey (1990) for analysis of Whalen's data because all models considered showed substantially better fit when this was done. Transformations of raw treatment values are routinely used to improve fit in probit analysis, where they are called "dose metameters" (Finney, 1971).

⁷The quasilielihood tests involve F ratios using deviance statistics from Table III and a heterogeneity coefficient estimated from Pearson X^2 statistics. The tests of increase of goodness-of-fit are analogous to "extra sum of square tests" in regression models. Some caution seems in order in interpreting the significance levels, since this approach makes rather restrictive assumptions about components of error that may not be appropriate for repeated measures data.

⁸Families of coefficients associated with the first column of Table IV are tested separately against the null hypothesis that each listener's coefficients are drawn from a multivariate normal population with a zero mean. The assumptions of these relatively simple tests appear to be plausible. However, newer, more complex techniques for analyzing generalized linear mixed models (GLMM's) may prove to be more powerful. Although GLMM's have received considerable attention in the recent statistics literature (e.g., Breslow and Clayton, 1993), no software using GLMM techniques capable of handling a problem the size of experiment I appears to be available.

⁹A conservative approach to the testing of multiple families employing Bonferroni correction would set the per-family significance level at 0.003 for an experiment-wise error rate of 0.05. By this conservative criterion, only the VCB interaction and the primary (VF, VS, CD, CB) and secondary cue (VD, FS) terms reach significance.

¹⁰The diphone bias effects are not discussed here. Diphone biases in model 6 (or any other model that includes $V \times C \times$ stimulus interactions) are not interpretable for the same reasons that group mean differences are not interpretable in analysis of covariance when there are separate slopes allowed for distinct groups (Snedecor and Cochran, 1967).

¹¹There may be some relation to Kingston and Diehl's (1995) suggestion that lower F_1 's can affect apparent gap duration when voice bar is present, but it is not clear to me how this would be reflected as VCD effects in the current analysis.

¹²Kingston and Diehl (personal communication) have provided me with data from an experiment they devised to replicate a subset of the stimuli described here. Analyses in both our laboratories agree that their experiment *does* show evidence for interaction patterns of the type they predict. The reason for the discrepancies is currently under investigation at both sites.

¹³There are a number of simplifying assumptions involved in the specific subclass of logistic regression models used here. If incorrect, these could result in lack of fit and could possibly lead to spurious diphone interactions analogous to what is sometimes called "removable nonadditivity" in ANOVA models. For example, the log of frequency and square root of duration transformations were chosen in advance of the analysis. Exploration of some other choices of metameter (see note 6) indicated that a linear time scale (ms) for voice bar duration leads to a slightly better overall fit for model 6. In this case, the VCB interaction, which showed the largest F -ratio of all the stimulus-tuned diphone terms in Table IV, was no longer significant by a random coefficients regression test [$F(4,11)=2.578$, $p<0.09$. In addition, though still significant, the F -ratios for the other two significant stimulus-by-diphone interactions were also decreased: for VCD, $F(4,11)=3.715$, $p<0.04$; for VCF, $F(4,11)=6.311$, $p<0.007$. The VCS interaction remained nonsignificant, $F(4,11)=1.481$, $p<0.27$].

¹⁴A set of output conditions sufficient to allow *optimal* recognition by diphone biased secondary cue models with linear boundaries [e.g., Fig. 3(c)] is discussed in Nearey (1992).

¹⁵Quadratic boundaries for perceptual data can be studied with logistic regression by including square and cross products of stimulus effects as additional variables.

Allen, J. (1994). "How do humans process and recognize speech?," IEEE Trans. Speech Audio Process. 2, 567-577.

Blumstein, S. (1986). "Comment on Lindau and Ladefoged," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 465-478.

Blumstein, S., and Stevens, K. (1980). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. 67, 648-662.

- Breslow, N. E., and Clayton, D. G. (1983). "Approximate inference in generalized linear mixed models," *J. Am. Stat. Assoc.* **88**, 9–25.
- Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English* (Harper, New York).
- Crowther, C., and Mann, V. (1992). "Native language factors affecting use of vocalic cues to final consonant voicing in English," *J. Acoust. Soc. Am.* **82**, 711–722.
- Diehl, R. L., and Kluender, K. R. (1989). "On the objects of speech perception," *Ecol. Psychol.* **1**, 121–144.
- Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).
- Finney, D. J. (1971). *Probit Analysis* (Cambridge U.P., Cambridge, England), 3rd. ed.
- Fisher, R. M., and Ohde, R. N. (1990). "Spectral and duration properties of front vowels as cues to final stop-consonant voicing," *J. Acoust. Soc. Am.* **88**, 1250–1259.
- Flége, J. E., Munro, M. J., and Fox, R. A. (1994). "Auditory and categorical effects on cross-language vowel perception," *J. Acoust. Soc. Am.* **95**, 3623–3641.
- Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. (1988). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," *J. Acoust. Soc. Am.* **84**, 115–123.
- Fowler, C. (1989). "Real objects of speech perception: a commentary on Diehl and Kluender," *Ecol. Psychol.* **1**, 145–169.
- Gumpertz, M., and Pantula, S. (1989). "A simple approach to inference in random coefficient models," *Am. Stat.* **43**, 203–210.
- Haberman, S. J. (1979). *Analysis of Qualitative Data, Volume 2* (Academic, New York).
- Jakobson, R., Fant, G., and Halle, M. (1963). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates* (MIT, Cambridge, MA).
- Jamieson, D. G., Nearey, T. M., and Ramji, K. (1989). "CSRE: a speech research environment," *Can. Acoust.* **17**, 23–25.
- Johnson, K., Ladefoged, P., and Lindau, M. (1993). "Individual differences in vowel production," *J. Acoust. Soc. Am.* **94**, 701–714.
- Kewley-Port, D., Watson, C., and Foyle, D. (1988). "Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli," *J. Acoust. Soc. Am.* **83**, 1133–1145.
- Kingston, J., and Diehl, R. (1994). "Phonetic knowledge," *Language* **70**, 419–454.
- Kingston, J., and Diehl, R. (1995). "Intermediate properties in the perception of distinctive feature values," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvanti (Cambridge U. P., Cambridge, England), pp. 7–27.
- Kingston, J., and Diehl, R. (1996). Personal communication.
- Klatt, D. (1981). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.
- Kohler, K. J. (1981). "Contrastive phonology and the acquisition of phonetic skills," *Phonetica* **38**, 213–226.
- Kohler, K. J. (1984). "Phonetic explanation in phonology: the feature fortis/lenis," *Phonetica* **41**, 150–174.
- Kuhl, P. (1992). "Infants' perception and representation of speech: development of a new theory," in *Proceedings ICSLP 92*, edited by J. Ohala, T. Nearey, B. Derwing, M. Hodge, and G. Wiebe (University of Alberta, Edmonton), pp. 449–455.
- Kuhl, P. K., 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.
- Labov, W. (1972). *Sociolinguistic Patterns* (University of Pennsylvania, Philadelphia).
- Ladefoged, P. (1990). "Some reflections on the IPA," *J. Phon.* **18**, 335–346.
- Ladefoged, P., DeClerk, J., Lindau, M., and Papçun, G. (1972). "An auditory-motor theory of speech production," *UCLA Working Papers in Phonetics* **22**, 48–75.
- Lahiri, A., Gewirth, L., and Blumstein, S. (1984). "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study," *J. Acoust. Soc. Am.* **76**, 391–404.
- Lieberman, A. M., and Mattingly, I. G. (1985). "The motor theory of speech perception revised," *Cognition* **21**, 1–36.
- Lindau, M., and Ladefoged, P. (1986). "Variability of feature specifications," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 465–478.
- Lindblom, B. (1986). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. Ohala and J. Jaeger (Academic Press, Orlando).
- Lindblom, B. (1990). "Explaining phonetic variation: a sketch of the H and H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Amsterdam), pp. 403–439.
- Martinet, A. (1955). *Économie des changements phonétiques* (A. Francke, Bern).
- Massaro, D., and Cohen, M. (1983). "Phonological context in speech perception," *Percept. Psychophys.* **34**, 338–348.
- Massaro, D., and Oden, G. (1980). "Evaluation and integration of acoustic features in speech perception," *J. Acoust. Soc. Am.* **67**, 996–1013.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models* (Chapman and Hall, London).
- McGowan, R. S., and Faber, A. (1996). "Introduction to papers on speech recognition and perception from an articulatory point of view," *J. Acoust. Soc. Am.* **99**, 1680–1682.
- Mermelstein, P. (1978). "On the relationship between vowel and consonant identification when cued by the same acoustic information," *Percept. Psychophys.* **23**, 331–335.
- Munro, M. J. (1992). "Perception and production of English vowels by native speakers of Arabic," Ph.D. thesis, University of Alberta.
- Nearey, T. M., (1980). "On the physical interpretation of vowel quality: cinefluorographic and acoustic evidence," *J. Phonetics* **8**, 213–241.
- Nearey, T. M. (1990). "The segment as a unit of speech perception," *J. Phon.* **18**, 347–373.
- Nearey, T. M. (1991). "Perception: Automatic and cognitive processes," in *Proceedings of the XII International Congress of Phonetic Sciences* (Publications de L'Université de Provence, Aix-en-Provence), Vol. 1, pp. 40–49.
- Nearey, T. M. (1992). "Context effects in a double-weak theory of speech perception," *Lang. Speech* **35**, 153–172.
- Nearey, T. M. (1995). "A double-weak view of trading relations: comments on Kingston and Diehl," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvanti (Cambridge U.P., Cambridge, England), pp. 28–40.
- Nearey, T., and Assmann, P. (1986). "Modeling the role of inherent spectral change in vowel identification," *J. Acoust. Soc. Am.* **80**, 1297–1308.
- Nearey, T., and Shammass, S. (1987). "Formant transitions as partly distinctive invariant properties in the identification of voiced stops," *Can. Acoust.* **15**, 17–24.
- Nossair, Z. B., and Zahorian, S. A. (1991). "Dynamic spectral shape features as acoustic correlates for initial stop consonants," *J. Acoust. Soc. Am.* **89**, 2978.
- Ohala, J. (1981). "The listener as a source of sound change," in *Papers from the Parasession on Language and Behavior*, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago Linguistic Society, Chicago), pp. 178–203.
- Ohala, J., and Shriberg, E. (1990). "Hypercorrection in speech perception," in *Proceedings of the 1990 International Conference on Spoken Language Processing* (Acoustical Society of Japan, Kobe), pp. 405–407.
- Perkell, J. S., Mathies, M. L., Svirsky, M. A., and Jordan, M. (1993). "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/," *J. Acoust. Soc. Am.* **93**, 2948–2961.
- Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," *J. Acoust. Soc. Am.* **32**, 693–703.
- Repp, B. (1982). "Phonetic trading relations and contexts effects: new evidence for a phonetic mode of perception," *Psychol. Bull.* **92**, 81–110.
- Scholes, R. (1967). "Phoneme categorization of synthetic vocalic stimuli by speakers of Japanese, Spanish, Persian and American English," *Lang. Speech* **10**, 46–68.
- Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods* (Iowa State U. P., Ames, IA).
- Stevens, K. (1990). "On the quantal nature of speech," *J. Phon.* **17**, 3–45.
- Stevens, K. N., and Blumstein, S. (1981). "The search for invariant acoustic correlates of phonetic features," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 1–38.
- Stevens, K. N., and Keyser, S. J. (1989). "Primary features and their enhancement in consonants," *Language* **65**, 81–106.
- Summers, W. V. (1987). "Effects of stress and final-consonant voicing on vowel production," *J. Acoust. Soc. Am.* **82**, 847–863.
- Summers, W. V. (1988). "F1 structure provides information for final-consonant voicing," *J. Acoust. Soc. Am.* **84**, 485–492.
- Sussman, H. M., Hoemeke, K. A., and Ahmed, F. S. (1993). "A cross-

- linguistic investigation of locus equations as a phonetic descriptor for place of articulation," *J. Acoust. Soc. Am.* **94**, 1256–1268.
- Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (1991). "An investigation of locus equations as a source of relational invariance for stop place categorization," *J. Acoust. Soc. Am.* **90**, 1256–1268.
- Sussman, H., Fruchter, D., and Cable, A. (1995). "Locus equations derived from compensatory articulation," *J. Acoust. Soc. Am.* **97**, 3112–3124.
- Watson, C. J., and Foyle, D. C. (1985). "Central factors in the discrimination and identification of complex sounds," *J. Acoust. Soc. Am.* **78**, 375–379.
- Watson, C., and Kewley-Port, D. (1988). "Some remarks on Pastore (1988)," *J. Acoust. Soc. Am.* **84**, 2266–2270.
- Whalen, D. (1989). "Vowel and consonant judgments are not independent when cued by the same information," *Percept. Psychophys.* **46**, 284–292.

Goupillaud inverse problem with arbitrary input

Wang Ning and Lin JunXuan

Ocean University of Qingdao, 5-Yushan Road, Qingdao 266003, People's Republic of China^{a)}

Ueha Sadayuki

Precision and Intelligent Laboratory, Tokyo Institute of Technology, 4259-Nagatada, Midori-Ku, Yokohama 227, Japan

(Received 29 June 1995; revised 15 January 1997; accepted 29 January 1997)

A new impulse inverse formula and its generalization including arbitrary input, are proposed. The new generalized inverse formula has the advantage of not being necessary to specify the first digital value of the input. Numerical instability of the inverse formula for arbitrary input is investigated carefully. An analysis of the perturbed stability and Lyapunov exponent are carried out. As the result of this analysis, it is shown that the estimated error is sensitive to the experimental error and previously estimated error as a sequential algorithm. Simulation using experimental data is also shown in the paper. © 1997 Acoustical Society of America. [S0001-4966(97)02306-0]

PACS numbers: 43.20.Dk, 43.20.Fn, 43.20.Ye [JEG]

INTRODUCTION

In the last decades, a large number of works have been contributed to the acoustic inverse scattering problem. The classical WKBJ solution was adopted by Hassab,¹ and Candel *et al.*,² to get an approximate solution for the acoustic inverse problem of one-dimensional layered medium. A similar technique was extended by Wang and Ueha to obtain an approximate solution of the problem inverting P , SV speed and density profiles with oblique incident scattering data.³ In addition to these approximate methods, exact inverse scattering method can be made for the one-dimensional inverse problem. Ware and Aki⁴ were the first ones to apply the formalism of the Gel'fand-Levitan method⁵ to inverse problem profiling the impedance from normally incidence impulsive reflection response. Berryman and Greene⁶ showed that there exists a close analogy between the Goupillaud inverse problem⁷ and the Gel'fand-Levitan inverse scattering problem for quantum physics. In fact, the Goupillaud inverse problem is a discrete version of inverse problem for the one-dimensional Schrödinger equation. There are other exact approaches for the same problem, among these there are time domain version of the Gel'fand-Levitan method discussed by Burridge,⁸ and a slightly different time domain formalism reported independently by Sondhi-Gopinath to make vocal tract shape identification.⁹ In the other hand, numerical method based on the characteristics is the most straightforward approach for the wave propagation analysis. The inverse problem methods based on this line were reported by Robinson,¹⁰ Claerbout,¹¹ and Bube and Burridge.¹² One of the advantages of the method using the characteristics over the exact approaches is that the former can be used for arbitrary input.

Although a lot of contributions have been made by many different people, two problems remain open. One is the problem to simultaneously invert the elastic P , SV speed and density profiles of layered elastic media; and another one is

numerical instability of inverse algorithm for arbitrary input. The first problem to reconstruct the elastic P , SV speed and density profile was investigated by Wang and Ueha,^{13,14} where the Marchenko method of inverse scattering problem for scalar waves has been extended to the problem for P - SV wave (vector wave). It was shown that, in principle one can reconstruct the profiles from the reflection and transmission coefficients (2×2 matrix). The second problem to be considered here, is the inverse algorithm behavior's instability for arbitrary input. Santosa and Schwetlick¹⁵ proposed a regularization method only for impulsive inverse problems except with increasing computing time. It will be seen that in fact the reasons of instability are twofold weak noise data may cause the numerical instability, on the other hand numerical behavior is also affected by properties of an input signal. To our knowledge, this problem has not been completely understood at the present time.

The Goupillaud inverse problem for plane waves will be considered in this paper. We shall generalize the inverse formula for impulse response to include the case of arbitrary input, and consider the problem of numerical instability.

I. GOUPILLAUD INVERSE PROBLEM AND A NEW INVERSE FORMULA

Consider the inverse scattering problem of one-dimensional layered medium, the medium is supposed to be a Goupillaud travel time layered medium, which is a number of homogeneous parallel layers of equal acoustic thickness (Fig. 1). Each interface at depth z_k is known by a reflection coefficient r_k , layer K lies between the interfaces $K-1$ and K with impedance $z_k = \rho_k C_k$. The goal of the inverse problem is to reconstruct reflectivity function or sequence $[r_0, r_1, r_2, \dots, r_N, \dots]$ (layer reflection coefficients) as a function of the travel time.

Ware and Aki⁴ have carried out an exact analysis for the impulsive inverse problem of the Goupillaud model to invert the impedance profile from impulsive reflection response. Their result can be easily extended to the case including arbitrary input. However, to apply this inverse formula to

^{a)}Please direct correspondence to this address.

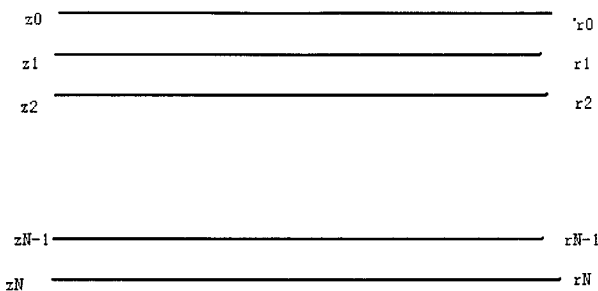


FIG. 1. The Goupillaud model of equal-travel-time layers, each layer k is characterized by its depth z_k and impedance $(\rho c)_k$

arbitrary input, one must precisely identify the first digital sampling value of incident waveform [see Eq. (1)]. This seems difficult for practical application, since the first digital sampling value of a practical incident wave generally only has poor signal-to-noise (S/N) ratio. To improve this problem, we derive a new inverse formula. Its generalization to arbitrary input can be performed without it being necessary to especially identify the first digital sampling value of input [Eq. (2)]. We give the lightface two inverse formulas below (a detailed derivation of these formula is left to Appendix A)

$$r_n = \frac{\sum_{i=0}^n f(n-1, i) R(n-i) - D(i) g(n-1, i-1)}{D(0) \prod_{i=0}^{n-1} (1-r_i^2)}, \quad (1)$$

and

$$r_n = \frac{\sum_{i=0}^n f(n-1, i) R(n-i) - D(i) g(n-1, i-1)}{\sum_{i=0}^n f(n-1, i) D(i) - R g(n-1, n-i+1)}. \quad (2)$$

Here $D(i)$ and $R(i)$ ($i=0,1,2,\dots$) are the data sequences of the incident and reflected signal, respectively. The functions $f(n, i)$ and $g(n, i)$ are completely determined by the recursive relation given in Appendix A. It is noted that the only difference between these two formulas is the denominator. These two formulas are the same but with different representation only.

The method presented here is sequential in the sense that $[r_0, r_1, r_2, \dots, r_v, \dots]$ the layer reflection coefficients are determined step by step from each new data of output time series and layer reflection coefficients previously estimated. However, it will be seen that numerical instability of inverse formula for arbitrary input is partly caused by this recursive processing. In what follows, we shall turn to consider this problem and show that the estimated error of the layer reflection coefficient is sensitive to experimental error. In particular, as a sequential algorithm, the time evolution of error is sensitive to previous data error.

As an example, we shall consider the error propagation of the formula (1) which is determined by

$$\Delta r_n = \frac{\Delta R(n) + \sum_{i=0}^{n-1} [\partial F_1(n) / \partial r(i)] \Delta r_i}{D(0) \prod_{i=0}^{n-1} (1-r_i^2)}, \quad (3)$$

where $\Delta R(i)$ ($i=1,2,\dots$) are the experiment data error, Δr_k are the estimated error of the layer reflection coefficients. $F_1(n)$ determined by (1) is a function of previous layer reflection coefficients r_k ($k=0,1,\dots,n-1$), its explicit form is given in Appendix B. It is noted that the error of the

k th layer comes from two parts: one is the experimental error $\Delta R(k)$, and another is the error from the estimated error of the previous layers. Hereafter, we shall always assume that the data series is normalized by the maximum value of the incident wave data.

At first, we rewrite the perturbed equation of Eq. (1) into a matrix form

$$M \Delta r = \Delta R, \quad (4)$$

where $\Delta R = [\Delta R(0), \Delta R(1), \dots, \Delta R(N)]$ and $\Delta r \equiv [\Delta r_0, \Delta r_1, \dots, \Delta r_N]$. The matrix M is a triangular one with diagonal elements being

$$\left[D(0), D(0)(1-r_0^2), \dots, D(0) \prod_{k=0}^{N-1} (1-r_k^2) \right], \quad (5)$$

where N is the total number of layers. (In practical application, N is taken as the total number of the sampling data.) The condition number of this matrix is proportional to $|D(0) \prod_{k=0}^{N-1} (1-r_k^2)|^{-1}$, so that the condition number becomes larger as N increases, and has an additional amplifying factor $[D(0)]^{-1}$ ($D(0) < 1.0$) with respect to the impulse inverse formula where $D(0) = 1$.

Next, to see that the time evolution of error is especially sensitive to previous data error, we investigate the time evolution of Δr_k providing that only the first digital sampling value $R(0)$ is contaminated with a small perturbation $\Delta R(0)$. After n iterations the error in r_k is generated according to the form of

$$\Delta r_n = \Delta R(0) \prod_{k=0}^{n-1} \frac{1}{D(0)^k} \left[\frac{F(k)}{\prod_{i=0}^{k-1} (1-r_i^2)} \right], \quad (6)$$

where explicit representation of $F(k)$ can be determined from (1) in the same way as F_1 . If a large number of iteration is performed, then Eq. (6) can be approximated by

$$\Delta r_n = \frac{[\langle T(n) \rangle_{av}]^n}{D(0)^n} \Delta R(0), \quad (7)$$

where $\langle T(n) \rangle_{av}$ is given by

$$\langle T(n) \rangle_{av} = \frac{\sum_{k=0}^{n-1} \log[F(k) / \prod_{i=0}^{k-1} (1-r_i^2)]}{n} \quad (8)$$

average value over n iterations. Thus, Eq. (7) can be recast to

$$\Delta r_n = e^{r \log[\langle T(n) \rangle_{av} / D(0)]} \Delta R(0), \quad (9)$$

then the corresponding Lyapunov exponent l_n is given by

$$l_n = \frac{1}{n} \log \frac{[\langle T(n) \rangle_{av}]}{D(0)}. \quad (10)$$

Equation (10) presents an approximate Lyapunov exponent. The exact one is given by taking the limit $n \rightarrow \infty$. However, it describes the average growth of errors. If the Lyapunov exponent is positive, the error in this discrete system will grow exponentially.

Now, it is clear that arbitrary input $D(0) (< 1.0)$ will amplify the condition number of the matrix M and the Lyapunov exponent l_n . We can conclude that the discrete system determined by (3) is more unstable than that for the

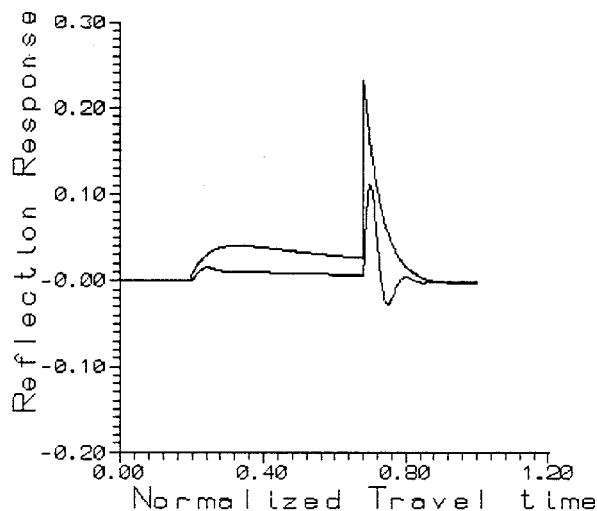


FIG. 2. Simulated reflection responses of the model in Fig. 1. Two different incident waveforms are used in simulation; one is an exponentially decaying wave; the other is a sinusoidally decaying wave.

impulsive case where $D(0)=1$. A similar argument can be made for Eq. (2) in the same way.

Remark:

It should be noted that whether an inverse algorithm is stable or not depends on both the properties of incident wave and the profile to be inverted, since $\langle T(n) \rangle_{av}$ is the function of both variables r_k ($k=0,1,\dots,N$) and $D(k)$ ($k=0,1,2,\dots,N$).

II. NUMERICAL SIMULATION

To verify the new formula, numerical simulations are performed to identify the impedance profile of the assumed model where the medium is assumed to be of 512 layers. The impulse response is generated by the following relation [from Eq. (1)],

$$R(n) = D(0)r_n \prod_{i=0}^{n-1} (1-r_i^2) - \sum_{i=0}^n f(n-1,i)R(n-i) + \sum_{i=0}^n g(n-1,i-1)D(i), \quad (11)$$

where r_n are the reflection coefficients of the n th layer. The first term in Eq. (11) means the direct reflection of the first incident wave data from the n th interface, and the remaining terms give the multireflection effects and the reflections induced by the successive incident wave data.

Numerical simulation for a gradiently varying impedance profile followed by an impedance jump is performed. Two different kinds of inlet pulse are used as the incident wave, the simulated reflection responses are shown in Fig. 2. From the simulated response, the layer reflection coefficients r_n are reconstructed by the formulas (1) and (2), and the impedance profiles are finally reconstructed from the relation

$$r_n = \frac{\rho_{n+1}c_{n+1} - \rho_n c_n}{\rho_{n+1}c_{n+1} + \rho_n c_n}. \quad (12)$$

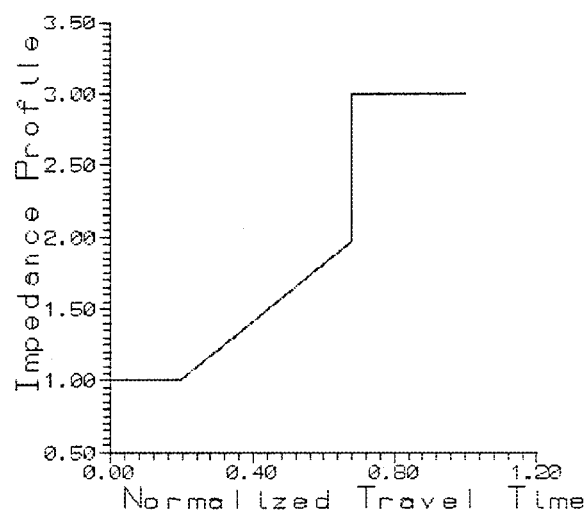


FIG. 3. Inverted impedance profile of the model for simulation. A gradiently varying region followed by an impedance jump.

In Fig. 3, the reconstructed impedance profiles are shown, the impedances inverted by different inverse formulas (1) and (2) are completely overlapped. As the result of this simulation, we can say that both the inverse formulas (1) and (2) behave well.

Next, we consider an inverse problem profiling the acoustic impedance using a practical experimental data. Figure 4 shows a typical pulse-echo signal in an experiment performed by Lin.¹⁶ The reconstructed impedance profile by (2) is shown in Fig. 5, where the result is divergent. The same calculation was performed by (1), but the result goes to overflow. In this numerical experiment, it seems to say that inverse formula (2) behaves better than formula (1). It may be the reason that for formula (2), one has an average value $D_{av}(0) [> D(0)]$ instead of $D(0)$ in Eq. (1), and this may stabilize the algorithm in the sense that the condition number

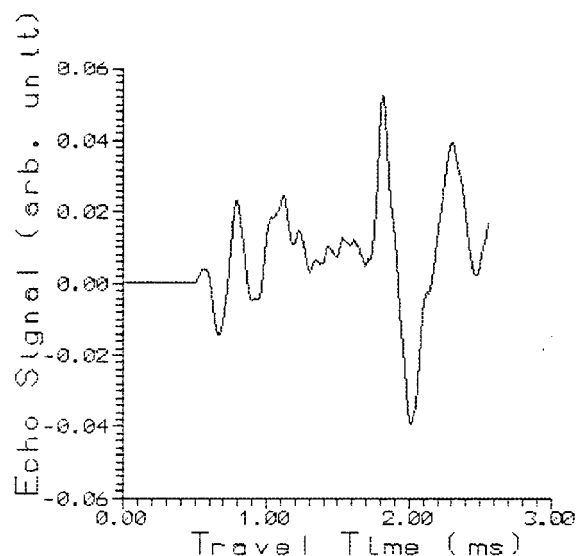


FIG. 4. A typical pulse-echo signal in the experiment. The medium in experiment consists of a gradiently varying sediment followed by the rock bottom.

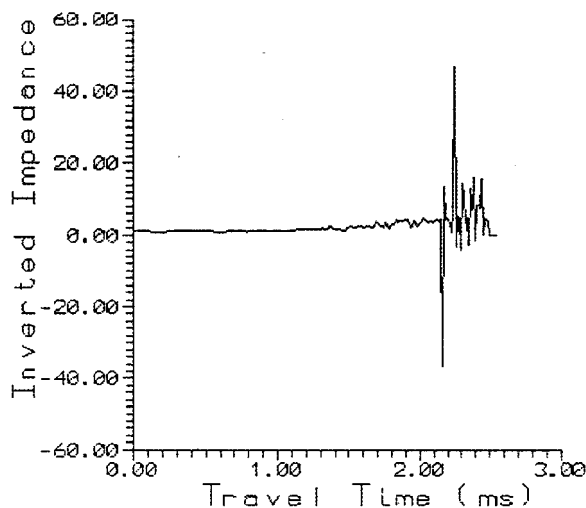


FIG. 5. Inverted impedance profile by using Eq. (2) without any preprocessing.

of the matrix M defined in Eq. (4) and the Laypunov exponent defined in Eq. (9) are decreased when $D_{av}(0) > D(0)$.

III. DECONVOLUTION OF IMPULSE RESPONSE

The numerical instability is a long standing problem affecting the practical application of the inverse scattering methods. The early research about the inverse scattering problem was focused on the impulse response inversion. In the past decade, a lot of attention has been given to the inverse scattering problem of arbitrary incident waveform. The only example where inverse formula for arbitrary input behaves well, is the recent report by Tenenbaum and Zindeluk.¹⁷ However, most of the inverse formula for arbitrary input is numerical instability. Nowadays, it is known that the inverse scattering problem is unstable in the presence of the noise data, and the numerical behavior is dependent on the properties of the incident waveform.¹⁸ These have been interpreted from view points of both simulation and theory presented in Sec. II. One of the results of our theoretical analysis is that the inverse scattering problem for impulse response behaves better than the general inverse scattering problem for arbitrary input. In a practical application, one usually extracts the impulse response from the practical signal as the first step, and then inverts the impedance profile by using a formula for impulse response inversion. Parra and Guerra¹⁸ estimated the impedance profile of layered sea sediment by a inversion scheme similar to the Goupillaud inversion scheme, where the impulse response is subtracted from the pulse-echo signal by using the frequency domain version of the Wiener filtering.¹⁹ The same deconvolution technique was also used by authors to invert the impedance profile of a functionally gradient material (FGM) sample.²⁰

The frequency domain version of the Wiener filtering used by those authors saves the computational time, but the efficiency of the filtering process obviously depends on a stabilizing constant, and in fact depends on prior knowledge about the signal-to-noise ratio. Here, we shall present a pre-filtering process using the LMS scheme. In what follows, the

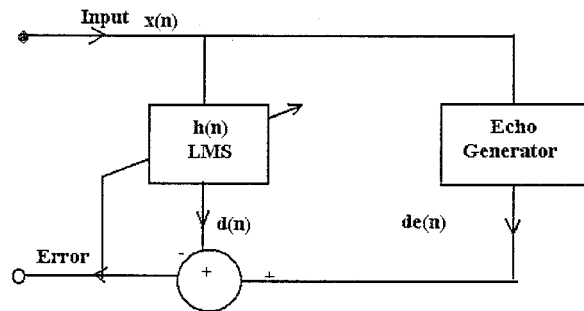


FIG. 6. Block diagram of the adaptive processor for extracting impulsive response.

echo generation is considered to be modeled as an N th-order finite impulse response by a transverse filter,

$$d(n) = \sum_{k=0}^N h(k)x(n-k), \quad (13)$$

where N is the total length of data, $x(t)$ is the data time series of the incident wave form, and $d(k)$ is the predicted echo series. We use the Widrow's LMS algorithm to minimize the cost (energy) function defined by²¹

$$E = \sum_n [d_{exp}(n) - d(n)]^2. \quad (14)$$

The impulse response is then extracted by canceling the unwanted incident wave. Figure 6 shows the flow diagram of this adaptive filter. The filtered time series of three different experimental data (sampled at different places) are used to invert impedance profiles. The resulted impedance profiles by the reduced inverse formula of Eq. (2) for impulsive response are shown in Fig. 7.

It has been measured by Lin that there exists a low speed layer at subsurface of the sediment, the corresponding impedance there is measured to be about $1.37 \times 10^6 \text{ Pa s/m}^2$ (see Ref. 16). The reconstructed impedance also shows a low

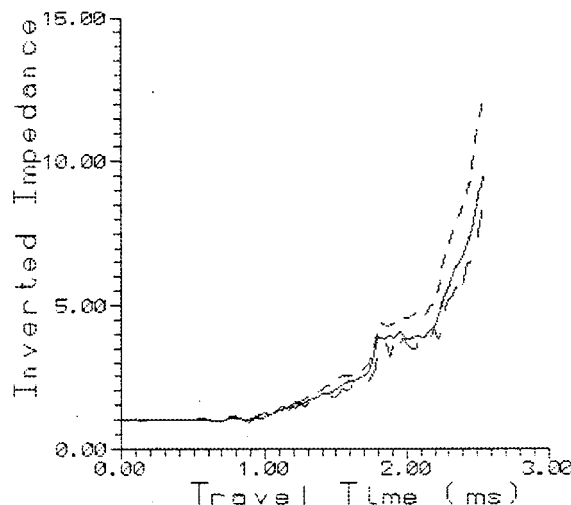


FIG. 7. Reconstructed impedance profiles by using the impulsive responses filtered by the adaptive processor. Three different profiles correspond to impedances at different sampling areas.

speed area, with the reconstructed impedance of corresponding low speed area being about $0.9 \times 1.5 (= 1.35) \times 10^6$ Pa s/m², and reproduce the gradient domain of impedance distribution and an impedance jump caused by the bottom of the oil pool.

IV. SUMMARY

The Goupillaud inverse formula was improved to include the case of an arbitrary input waveform, and a new impulsive inverse formula and its generalization for any input was derived. A careful investigation of the numerical instability of inverse formula for any input was carried out, which interprets why the inverse formula for any input behaves more poorly than that of the impulsive one. The reason is that the arbitrary input wave data will increase the Lyapunov exponent, and make it positive. To profile a practical impedance, the impulse response was filtered by using the Widrow's LMS algorithm. Both the two impulsive inverse formulas were used to invert the impedance and behaviors well for practical experimental data.

ACKNOWLEDGMENT

This work was partly supported by the NSFC of China under Grant Nos. 19234040 and 19474028.

APPENDIX A

In this Appendix, we should give a sketch of the derivation of the new impulse inverse formula proposed by authors recently. Following Ware and Aki (see Ref. 4), the reflection coefficients $R(z)$ ($z = e^{2i\omega\Delta}$, Δ =time sampling) for arbitrary input is determined by the relation,

$$R(z) = \frac{z^{n+1}G(n, 1/z)}{F(n, z)} D(z), \quad (A1)$$

where $F(n-1, z)$ and $G(n-1, z)$ are defined by

$$\prod_{i=0}^n t_i F(n, z) = \sum_{i=0}^n f(n, i) z^i, \quad (A2)$$

$$\prod_{i=0}^n t_i G(n, z) = \sum_{i=0}^n g(n, i) z^i,$$

and $f(n, i)$ and $g(n, i)$ is determined recursively by

$$\begin{aligned} f(n+1, i) &= f(n, i) + r_{n+1}g(n, i-1), \\ g(n+1, i) &= g(n, i-1) + r_{n+1}f(n, i). \end{aligned} \quad (A3)$$

Using (A3) the following recursive relation for $F(n, z)$ and $G(n, z)$ can be obtained easily,

$$\begin{aligned} t_n F(n, z) &= F(n-1, z) + r_n z G(n-1, z), \\ t_n G(n, z) &= z G(n-1, z) + r_n F(n-1, z) \end{aligned} \quad (A3')$$

combining Eqs. (A2) and (A3) or (A3'), Eq. (A1) can be rewritten to

$$r_n \left\{ \sum_{N=2}^{\infty} z^N \sum_{i=0}^{N-1} f(n-1, n+1+i-N) D(i) \right.$$

$$\begin{aligned} &- z \sum_{N=1}^{\infty} z^N \left[\sum_{i=0}^{N-1} R(i) g(n-1, N-i-1) \right] \\ &= - \left\{ \sum_{N=1}^{\infty} \left[z^N \sum_{i=0}^{n-1} D(i) g(n-1, n+i-N) \right] \right. \\ &\quad \left. + \sum_N z^N \left[\sum_{i=0}^{N-1} R(i) f(n-1, N-i-1) \right] \right\}. \end{aligned} \quad (A4)$$

Now, comparing the coefficients of power z^{n+1} at both sides of Eq. (A4) yields

$$r_n = \frac{\sum_{i=0}^n f(n-1, i) R(n-i) - D(i) g(n-1, i-1)}{\sum_{i=0}^n f(n-1, i) D(i) - R(i) g(n-1, n-i+1)}. \quad (A5)$$

The impulsive inverse formula can be obtained by virtue of the fact $D(0)=1$, $D(i)=0$ otherwise for impulse incidence and $f(n-1, 0)=1$,

$$r_n = \frac{\sum_{i=0}^n f(n-1, i) R(n-i)}{1 - \sum_{i=0}^{n-1} R(i) g(n-1, n-i+1)}. \quad (A6)$$

For comparison, we also give out the original formula of Ware-Aki

$$r_n = \frac{\sum_{i=0}^n R(i) f(n-1, n-i)}{\prod_{i=0}^{n-1} (1 - r_i^2)}. \quad (A7)$$

APPENDIX B

A straightforward calculation yields

$$\begin{aligned} \frac{\partial F_1}{\partial r_i} &= - \frac{2r_i}{(1-r_i^2)} \left\{ \sum_{k=1}^{n-1} f(n-1, k) R(n-k) - D(k) \right. \\ &\quad \times g(n-1, k-1) \left. \right\} + \left\{ \sum_{k=1}^{n-1} \frac{\partial}{\partial r_i} f(n-1, k) R(n-k) \right. \\ &\quad \left. - D(k) \frac{\partial}{\partial r_i} g(n-1, k-1) \right\}, \end{aligned} \quad (B1)$$

where $(\partial/\partial r_i)f(n-1, k)$, $(\partial/\partial r_i)g(n-1, k)$ can be determined recursively by (A3).

- ¹J. C. Hassab, "A generalized approach to the solution of variable systems subjected an arbitrary source function and boundary conditions," J. Sound Vib. **48**, 277-291 (1976).
- ²S. M. Candel, F. De Fillipi, and A. Launay, "Determination of the inhomogenous structure of a medium from its plane wave reflection response. Part I. A numerical analysis of direct problem," J. Sound Vib. **68**, 571-582 (1980).
- ³N. Wang and S. Ueha, "Direct Scattering and WKB-Born inverse problem for a layered elastic medium using the First-Order equations of P-SV motion," Acustica **73**, 248-253 (1991).
- ⁴A. K. Ware and K. Aki, "Continuous and discrete inverse scattering problem in stratified elastic medium: I: Plane waves at normal incidence," J. Acoust. Soc. Am. **45**, 911-921 (1969).
- ⁵I. M. Gel'fand and B. M. Levitan, "On the determination of a different equation from the spectral function," Am. Math. Soc. Transl. **1**(2), 253-304 (1955).
- ⁶J. G. Berryman and R. R. Greene, "Discrete inverse methods for elastic waves in layered media," Geophysics **45**(2), 213-232 (1980).
- ⁷P. L. Goupillaud, "An approach to inverse filtering of near-surface layer effects from seismic records," Geophysics **26**, 354 (1961).
- ⁸R. Burridge, "The Gel'fand-Levitan, the Marchenko and Gopinath-Sondhi integral equation of inverse scattering theory, regarding the con-

- text of inverse impulse response problems," *Wave Motion* **2**, 305–325 (1980).
- ⁹M. M. Sondhi and B. Gopinath, "Determination of vocal tract shape from impulse response at the lips," *J. Acoust. Soc. Am.* **48**, 1869–1873 (1971).
 - ¹⁰E. A. Robinson and S. Treitel, *Geophysical Signal Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
 - ¹¹J. Claebout, *Fundamentals of Geophysical Data Processing* (McGraw-Hill, New York, 1976).
 - ¹²K. P. Bube and R. Burridge, "The one-dimensional inverse problem of reflection seismology," *SIAM (Soc. Ind. Appl. Math.) Rev.* **25**, 497 (1983).
 - ¹³N. Wang and S. Ueha, "Scattering of elastic waves in stratified isotropic media. II. Inverse problem involving P and SV waves," *J. Acoust. Soc. Jpn.* **45**, 845–854 (1989), in Japanese.
 - ¹⁴N. Wang and S. Ueha, "Inverse Scattering of a medium which supports M -types of wave," *J. Phys. Soc. Japan* **61**, 455–461 (1992).
 - ¹⁵F. Santosa and H. Schwetlick, "The inversion of acoustical impedance profile by methods of characteristics," *Geophys. J. R. Astron. Soc.* **70**, 229–243 (1982).
 - ¹⁶J. X. Lin, "Verification of the stability criterion of inverse algorithms for layered sea bottom," *Chin. J. Acoust.* **12**, 84–88 (1993).
 - ¹⁷R. A. Tenenbaum and M. Zindeluk, "An exact solution for the one-dimensional elastic wave equation in layered media," *J. Acoust. Soc. Am.* **92**, 3364–3378 (1992).
 - ¹⁸A. Bamberger and C. Chavent, "About the stability of the inverse problem in 1-D wave equations—Application to the interpretation of seismic profiles," *Appl. Math. Opt.* **5**, 163–174 (1977).
 - ¹⁹P. C. Parra and C. R. Guerra, "Impedance profile and overall attenuation estimation of layered sea bottoms from their normal incident acoustic reflection response," *J. Acoust. Soc. Am.* **85**, 2388–2393 (1989).
 - ²⁰O. Kuwahara, N. Wang, and S. Ueha, "Reconstruction of Acoustic impedance distribution of functionally gradient materials by using reflection impulse response," *Jpn. J. Appl. Phys.* **1** **31**, 102–104 (1991).
 - ²¹S. T. Alexander, *Adaptive Signal Processing, Theory and Application* (Springer-Verlag, New York, 1986).

On the scattering of antisymmetric edge modes

M. de Billy

Groupe de Physique des Solides, Universités Paris 6 et 7, 2 place Jussieu, 75251, Paris Cedex 05, France

(Received 18 November 1996; accepted for publication 25 February 1997)

The scattering of antisymmetric flexural waves is illustrated via three series of experiments conducted on linear elastic solid wedges immersed in air. It is quantitatively demonstrated that scattering into different modes exists when an antisymmetric mode is reflected at the end face of a wedge or when it is submitted to an external pressure applied on the tip of the wedge. The existence of transmission of edge modes is also experimentally confirmed. © 1997 Acoustical Society of America. [S0001-4966(97)04706-1]

PACS numbers: 43.20.Fn [ANN]

INTRODUCTION

In the recent past antisymmetric flexural modes have been extensively studied, theoretically and experimentally. It was confirmed¹⁻⁹ that the number of modes and the velocity of these acoustic waveguides vary with the value of the apex angle of the wedge. Most of the investigations report results on the propagation of these antisymmetric modes and very few^{10,11} are concerned with the interaction of these waves with discontinuity. The processes of multimode wedge wave scattering on small imperfections were for the first time investigated theoretically by Krylov and Shanin.¹¹ The main difference of the present work from the earlier published experimental investigation of wedge wave scattering¹⁰ is that, considering slender wedges, it deals with a more general case of multimode propagation and conversion.

The first objective of the present experimental study is the analysis of the influence on the propagation of ASF acoustic modes of a local damping exerted on the tip of a wedge. This type of experiment will be named hereafter “forward scattering” experiments. The second objective of this work is concerned with the study of the “transmission” of ASF modes through an interface between coupled wedges of different materials and various apex angles. The last purpose deals with the “backscattering” (or reflection) of an incident edge wave by the free end face of a wedge.

The time wavetrains obtained by transmission (two transducers) or by reflection (one transducer) point out the existence of anomalies in the echographic structure and the experimental results demonstrate that the additional echoes are caused by the conversion of the incident wave into ASF modes of higher and lower order.

I. EXPERIMENTAL CONDITIONS

Two series of experiments are described. The first one includes the forward scattering and the transmission measurements. The edge waves are acoustically generated by the technique described in Refs. 10 and 12 and initiated by Pajewski *et al.*¹³ It uses two shear contact transducers coupled with the extremities *A* and *B* of the sample [see Fig. 1(a) and (b)]; In case of forward scattering experiments, the local pressure is realized with a sticky paste [point *G* in Fig. 1(a)] stuck on the linear edge and on the two planes of the wedge over a surface of about 5×5 mm². The second series of

experiments deals with reflection measurements. In that case, the opposite end face (*B*) of the wedge is free [see Fig. 1(c)] and only one shear contact transducer acts as the source and the receiver. In both cases, the emitter is driven with short pulses of 1 MHz in central frequency. The direction of the polarization of the transmitters is perpendicular to the bisector plane for the generation and the detection of the ASF modes.^{12,13} Usually the coupling area (*s*) between the transmitters and the sample is kept small to eliminate the undesired bulk and surface waves.¹² The experiments are realized at room temperature and on acute linear wedges.

All the measurements realized in forward and backward geometries were achieved on duraluminum wedges of 400.0±0.1 mm in length (the length of the wedge will be named *L* hereafter). The apex angle (Φ) ranges from 30° to 80°. For transmission analysis, shorter samples (*L*=70.0±0.1 mm) of different materials (Polyvinyl, Plexiglas, and duraluminum) and of only two apex angles (35° and 65°)

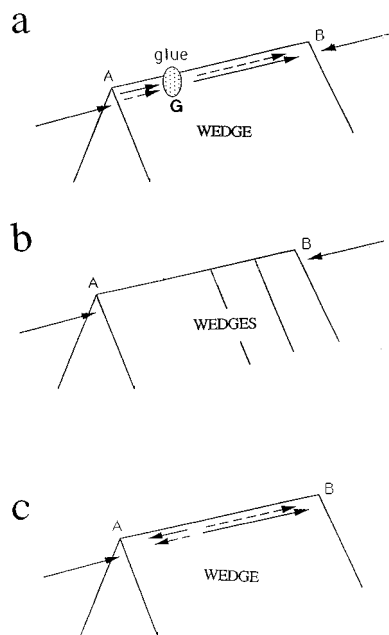


FIG. 1. Definition of the geometrical arrangements: (a) case of the “pseudo-transmission”; (b) case of the “true-transmission”; (c) case of the reflection.

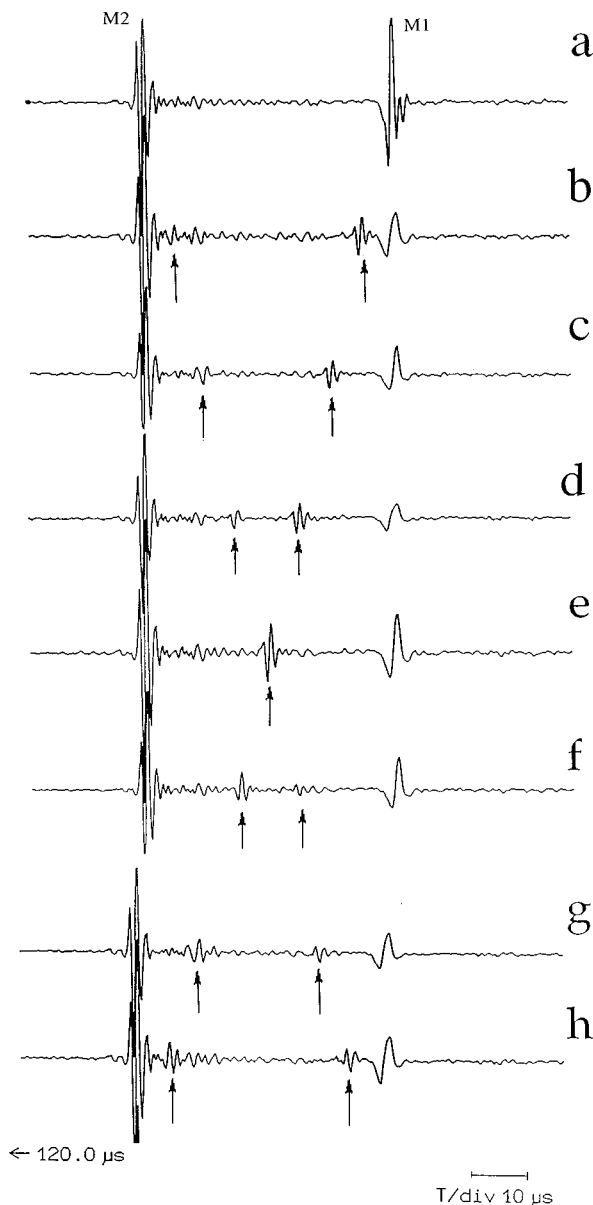


FIG. 2. Wavetrains obtained for different positions of the glue (case of a 50° apex angle duraluminum wedge; $AB=400$ mm). (a) Free sample; (b) $d=50$ mm; (c) $d=100$ mm; (d) $d=150$ mm; (e) $d=200$ mm; (f) $d=250$ mm; (g) $d=300$ mm; (h) $d=350$ mm.

were submitted to investigation. The samples were machined as sharp as possible in order to avoid the effects of dispersion due to the truncation.^{5,6,14}

II. FORWARD SCATTERING MEASUREMENTS

We present in this section results obtained with forward scattering experiments which are limited to the antisymmetric modes propagating in linear elastic wedges.

A. Description of the wavetrains

The recorded wavetrains obtained with a linear duraluminum wedge of 50° apex angle are presented in Fig. 2 for different positions of the sticky strip on the tip of the wedge. For comparison, the wavetrain redrawn in Fig. 2(a) is char-

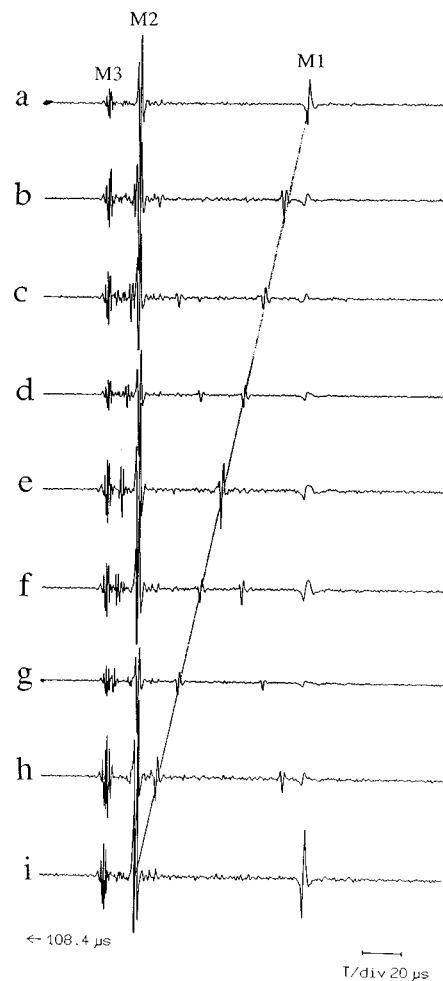


FIG. 3. Wavetrains recorded for different positions of the glue (case of a 35° apex angle duraluminum wedge; $AB=400$ mm). (a) Free sample; (b) $d=50$ mm; (c) $d=100$ mm; (d) $d=150$ mm; (e) $d=200$ mm; (f) $d=250$ mm; (g) $d=300$ mm; (h) $d=350$ mm; (i) $d=400$ mm.

acteristic of a sample without any glue on the tip of the wedge. Two ASF modes ($M1$ and $M2$) are detected.

It is interesting to notice that when the glue is present [Fig. 2(b)–(h)] two echoes exist (indicated by an arrow) in addition to those obtained when the tip of the wedge is free [Fig. 2(a)]. As the distance d between the origin (A) of the edge and the middle (G) of the glued space increases, these echoes move toward each other [Fig. 2(b)–(d)]. When d is equal to the half-length of the sample, the two signals interfere constructively [Fig. 2(e)] and give rise to a larger echo. After they cross, a behavior symmetric to the one observed for $d < L/2$ is noticed [Fig. 2(f)–(h)].

Another example of these additional echoes is given in Fig. 3; it depicts the evolution of the temporal signals recorded by the receiver for different positions of the glue in the case of a 35° apex angle duraluminum sample ($L=400.0 \pm 0.1$ mm). For this value of the apex angle, a third echo is recorded in addition to the two first ASF modes [Fig. 3(a)]; it is due to the third ASF mode ($M3$). Between modes $M1$ and $M2$, it is again verified that two echoes move linearly with $d(=AG)$ in two opposite directions [Fig. 3(b)–(h)]: One comes from the echo which corresponds to the

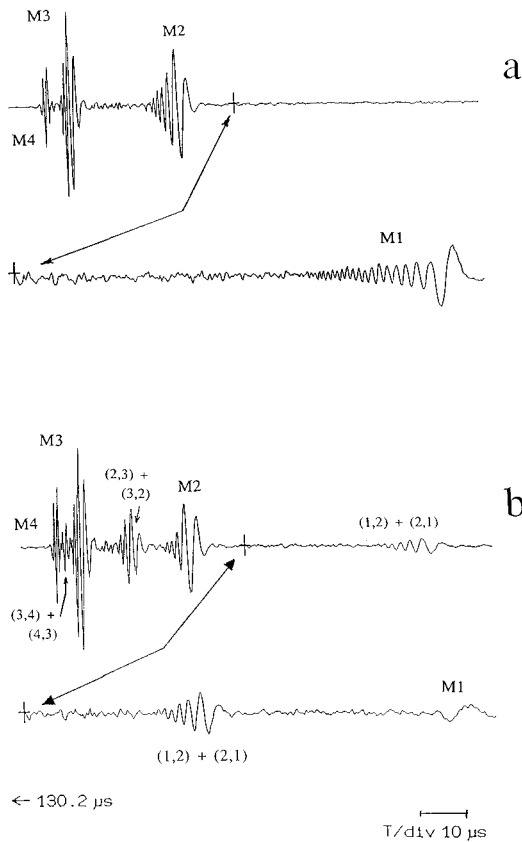


FIG. 4. Wavetrains recorded with a 30° apex angle duraluminum wedge: (a) without glue; (b) with glue on $L/2$.

fundamental ASF mode (mode $M1$) and the other one moves away from the echo which is identified as the second ASF mode (mode $M2$). A similar behavior is observed between the second mode and the mode $M3$. However, because of the modes $M2$ and $M3$ are close in the time domain, the phenomenon is clear only in the case for which d is approximately equal to the half-length of the wedge [Fig. 3(e)].

In the previous records, the signals were not dispersive. For wavetrains showing a dispersive behavior, the same phenomenon is observed. This is clearly illustrated in Fig. 4 which compares the wavetrains recorded without glue [Fig. 4(a)] and with glue positioned on $d=L/2$ [Fig. 4(b)] for a duraluminum sample ($\Phi=30^\circ$). In the first picture, three

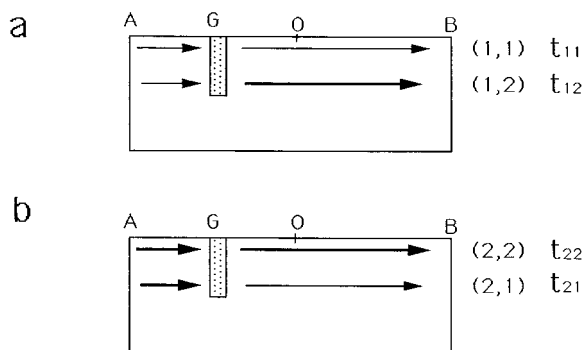


FIG. 5. Schematic diagram of the identification of the modes ($d < L/2$).

TABLE I. Values of the velocity of the different modes obtained from Ref. 12 and used for calculation of the arrival times. These values are obtained within an accuracy better than 0.5%.

| Φ -Material | $M1$ | Mode $M2$ | $M3$ |
|------------------|------|-----------|------|
| 35°-duraluminum | 1649 | 2504 | 2805 |
| 45°-duraluminum | 1957 | 2791 | ... |
| 50°-duraluminum | 2187 | 2820 | ... |
| 65°-duraluminum | 2503 | ... | ... |
| 35°-Plexiglas | 748 | 1179 | ... |
| 65°-Plexiglas | 1113 | ... | ... |
| 35°-Polyvinyl | 582 | 914 | ... |
| 65°-Polyvinyl | 865 | ... | ... |

ASF modes ($M1$, $M2$, and $M3$) are identified without any ambiguity. As mentioned before, the earliest echo may be associated to the ASF mode ($M4$). The temporal record shown in Fig. 4(b) is characterized by the existence of three new echoes positioned midway between the signals which correspond to the excited guided modes.

B. Analysis of the experimental records

The experimental records described in the previous section and concerned by the effect of a damping on the propagation of the ASF modes in a linear elastic wedge may be justified by considering a scattering process giving rise to mode conversions at the position where the glue is stuck. The incident mode M_i is converted into a faster mode $M_{j=i+1}$ and a slower mode $M_{j=i-1}$ (with $i > 2$).

Consider the geometrical situation schematically described in Fig. 5. The point O is such that $2AO=AB$ and the distance AG verifies the inequality $AG < AO = L/2$. Two ASF modes $M1$ and $M2$ are supposed to be generated in A . If a process of mode conversion is considered in G , the incident mode $M1$ [Fig. 5(a)] either goes through G —mode (1,1)—or is converted into a mode $M2$ —mode (1,2); in this nomenclature (i,j), the first index ($i=1,2,\dots$) identifies the incident mode and the second index the “pseudo-transmitted” ASF mode ($j=1,2,\dots$). Similarly the modes (2,2) and (2,1) are described in Fig. 5(b). The calculated values of the arrival times (t_{ij}) of the different echoes are given by the following equations in which C_1 and C_2 are

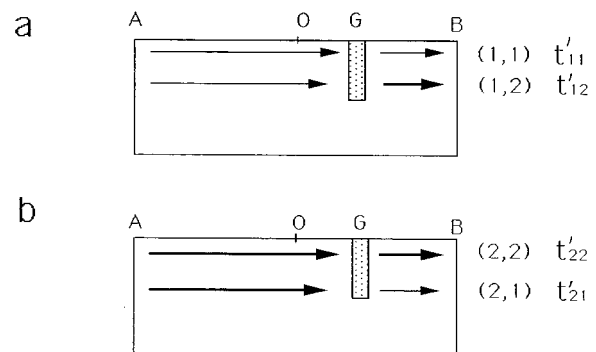


FIG. 6. Schematic diagram of the identification of the modes ($d > L/2$).

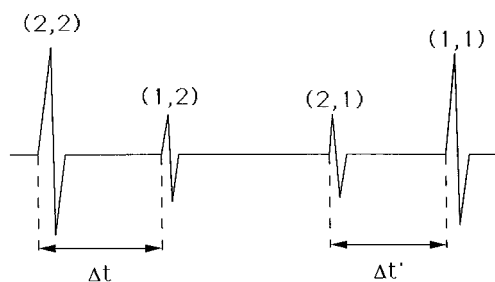


FIG. 7. Schematic diagram showing the symmetrical structure of the wavetrains due to the processus of conversion mode.

the velocity of propagation of the modes $M1$ and $M2$, respectively, for the investigated wedge:

$$\begin{aligned} \text{mode (1,1)} \quad t_{11} &= L/C_1 \\ \text{mode (1,2)} \quad t_{12} &= d/C_1 + (L-d)/C_2 \\ \text{mode (2,2)} \quad t_{22} &= L/C_2 \\ \text{mode (2,1)} \quad t_{21} &= d/C_2 + (L-d)/C_1. \end{aligned} \quad (1)$$

The values of C_i used for calculations are obtained from the experimental data in Ref. 12 and are reported in Table I. These experimental data are obtained with a relative error

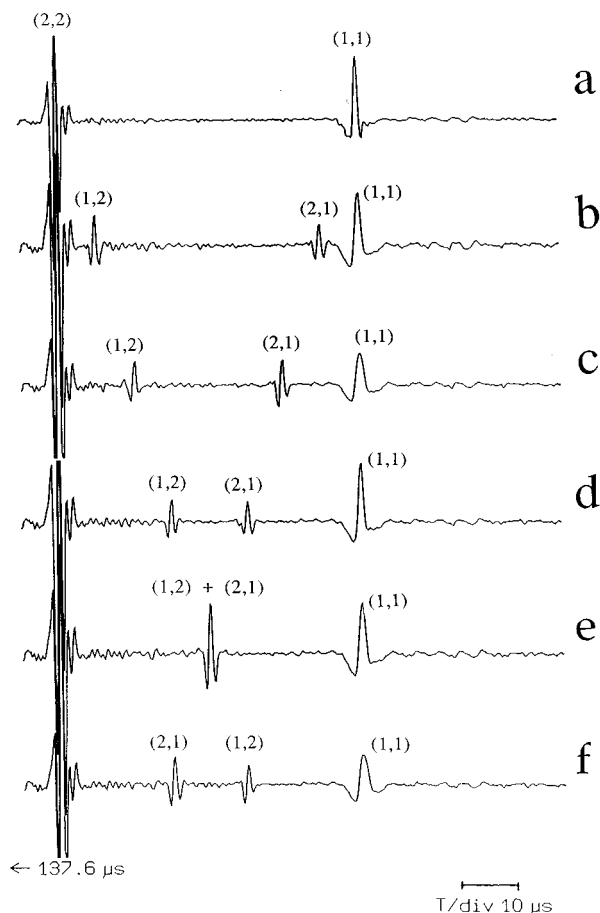


FIG. 8. Pseudo-transmitted wavetrains obtained for different positions of the glue (case of a 45° apex angle duraluminum wedge; $L=400$ mm): (a) $d=0$; (b) $d=50$ mm; (c) $d=100$ mm; (d) $d=150$ mm; (e) $d=200$ mm; (f) $d=250$ mm.

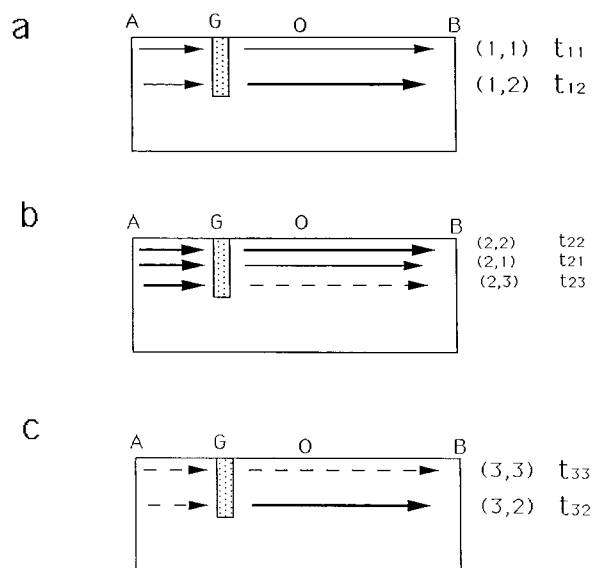


FIG. 9. Schematic diagrams illustrating the conversion modes in case of three incident ASF modes.

better than 0.5%. As C_1 is lower than C_2 , the calculated arrival times verify the inequalities: $t_{11} > t_{21} > t_{12} > t_{22}$ for the particular investigated geometrical configuration ($d < L/2$). The same data may be calculated when the glue is positioned at a distance d such that $d > L/2$ (Fig. 6). If the arrival times are t'_{ij} , it is easy to verify that $t_{ii} = t'_{ii}$ and $t_{ij} = t'_{ij}$. This relation induces that the time differences Δt and $\Delta t'$ are equal (cf. Fig. 7):

$$\Delta t = t_{12} - t_{21} = d(1/C_1 - 1/C_2) = \Delta t'.$$

This result confirms that the wavetrain, experimentally recorded, presents a symmetrical structure with respect to the

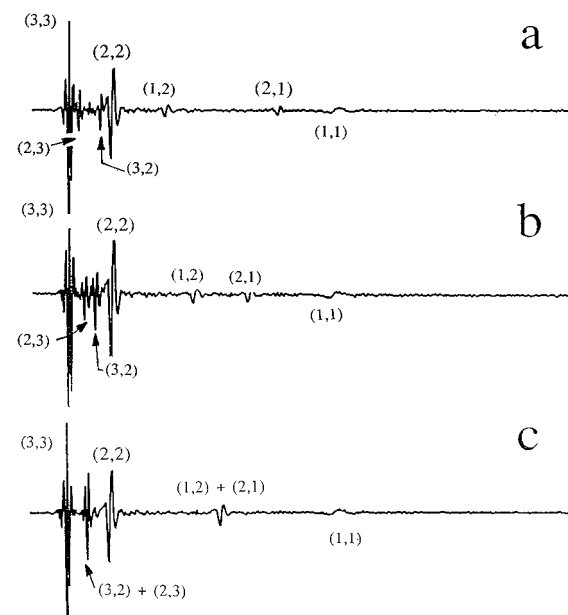


FIG. 10. Pseudo-transmitted wavetrains obtained for different positions of the glue (case of a 35° apex angle duraluminum wedge; $L=400$ mm): (a) $d=100$ mm; (b) $d=150$ mm; (c) $d=200$ mm.

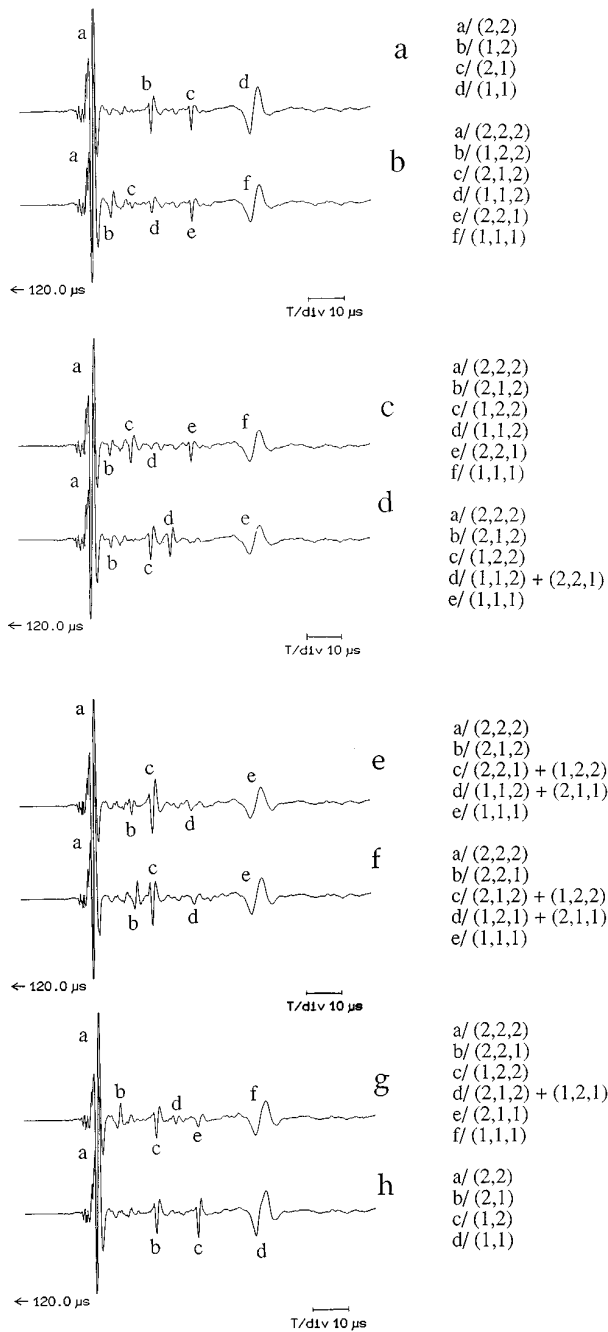


FIG. 11. Identification of the different modes.

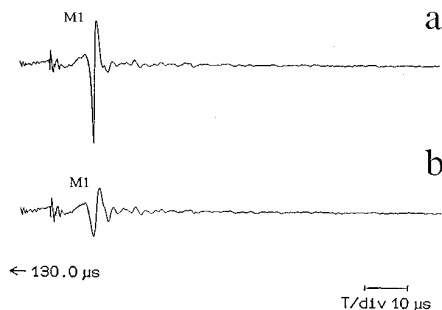


FIG. 12. Waveforms recorded with a 70° apex angle duraluminum wedge: (a) without glue; (b) with glue on $L/2$.

TABLE II. Comparison between the calculated and the measured values of the arrival time of the modes (1,2) and (2,1) for different values of the distance d .

| Distance d (in mm) | Arrival time (μ s) of the mode (1,2) | | Arrival time (μ s) of the mode (2,1) | | Ref. (Fig. 13) |
|-------------------------|--|-----------------|--|-----------------|-------------------|
| | Calculated | Measured | Calculated | Measured | |
| 50 | 150.0 ± 1.5 | 149.9 ± 0.5 | 191.4 ± 1.8 | 192.2 ± 0.3 | b |
| 100 | 157.1 ± 1.3 | 156.2 ± 0.5 | 184.6 ± 1.4 | 184.4 ± 0.3 | c |
| 150 | 163.9 ± 1.1 | 164.2 ± 0.5 | 177.7 ± 1.2 | 178.2 ± 0.3 | d |
| 200 | 170.8 ± 1.0 | 171.2 ± 0.4 | 170.8 ± 1.0 | 171.2 ± 0.4 | e |
| 250 | 177.7 ± 1.2 | 178.2 ± 0.4 | 157.1 ± 1.1 | 164.2 ± 0.4 | f |

TABLE III. Comparison between the calculated and the measured values of the arrival time of the modes (1,2), (2,1), (2,3), and (3,2) for different values of the distance d .

| Distance d (in mm) | Arrival time (μ s) of the mode (1,2) | | Arrival time (μ s) of the mode (2,1) | | |
|-------------------------|--|-----------------|--|-----------------|--|
| | Calculated | Measured | Calculated | Measured | |
| 100 | 180.9 ± 1.4 | 179.0 ± 0.8 | 222.9 ± 2.1 | 221.0 ± 0.9 | |
| 150 | 191.4 ± 1.3 | 189.7 ± 0.9 | 212.4 ± 1.7 | 210.4 ± 1.0 | |
| 200 | 201.9 ± 1.2 | 201.0 ± 1.0 | 201.9 ± 1.2 | 201.0 ± 1.0 | |

| Distance d (in mm) | Arrival time (μ s) of the mode (2,3) | | Arrival time (μ s) of the mode (3,2) | | |
|-------------------------|--|-----------------|--|-----------------|--|
| | Calculated | Measured | Calculated | Measured | |
| 100 | 147.1 ± 1.1 | 145.7 ± 0.7 | 155.7 ± 1.3 | 154.2 ± 0.8 | |
| 150 | 149.3 ± 1.3 | 147.7 ± 0.7 | 153.6 ± 1.3 | 153.1 ± 0.8 | |
| 200 | 151.4 ± 1.4 | 150.4 ± 0.8 | 151.4 ± 1.4 | 150.4 ± 0.8 | |

TABLE IV. Description of the different investigated geometries.

| Distance (in mm) | | | |
|------------------|--------|----------|--------|
| curve | AG_1 | G_1G_2 | G_2B |
| 11-(a) | 150 | 0 | 250 |
| 11-(b) | 50 | 100 | 250 |
| 11-(c) | 100 | 50 | 250 |
| 11-(d) | 150 | 50 | 200 |
| 11-(e) | 150 | 100 | 150 |
| 11-(f) | 150 | 150 | 100 |
| 11-(g) | 150 | 200 | 50 |
| 11-(h) | 250 | 0 | 150 |

TABLE V. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (a)].

| Mode | Calculated value in μ s | Measured value in μ s |
|-------|--------------------------------|------------------------------|
| (2,2) | 141.8 ± 0.9 | 139.0 ± 4.0 |
| (1,2) | 157.2 ± 1.0 | 156.1 ± 1.5 |
| (2,1) | 167.5 ± 1.1 | 168.0 ± 0.5 |
| (1,1) | 182.9 ± 1.2 | 183.3 ± 1.8 |

TABLE VI. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (b)].

| Mode | Calculated value in μs | Measured value in μs |
|---------|-----------------------------------|---------------------------------|
| (2,2,2) | 141.8 ± 1.7 | 139.2 ± 4.1 |
| (1,2,2) | 146.9 ± 1.7 | 145.0 ± 0.8 |
| (2,1,2) | 150.9 ± 1.8 | 151.0 ± 0.4 |
| (1,1,2) | 157.2 ± 1.9 | 156.6 ± 0.6 |
| (2,2,1) | 167.5 ± 2.0 | 168.0 ± 0.2 |
| (1,2,1) | 172.5 ± 2.0 | ... |
| (2,1,1) | 177.7 ± 2.1 | ... |
| (1,1,1) | 182.9 ± 2.1 | 183.3 ± 1.8 |

TABLE VII. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (c)].

| Mode | Calculated value in μs | Measured value in μs |
|---------|-----------------------------------|---------------------------------|
| (2,2,2) | 141.8 ± 1.7 | 139.0 ± 4.0 |
| (2,1,2) | 146.9 ± 1.7 | 145.1 ± 1.4 |
| (1,2,2) | 150.9 ± 1.8 | 151.3 ± 3.0 |
| (1,1,2) | 157.2 ± 1.9 | 157.3 ± 0.6 |
| (2,2,1) | 167.5 ± 2.0 | 168.0 ± 0.5 |
| (2,1,1) | 172.5 ± 2.1 | ... |
| (1,2,1) | 177.7 ± 2.1 | ... |
| (1,1,1) | 182.9 ± 2.1 | 183.3 ± 1.7 |

TABLE VIII. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (d)].

| Mode | Calculated value in μs | Measured value in μs |
|-------------------|-----------------------------------|---------------------------------|
| (2,2,2) | 141.8 ± 1.7 | 139.0 ± 4.0 |
| (2,1,2) | 146.9 ± 1.7 | 147.2 ± 0.4 |
| (1,2,2) | 157.2 ± 1.9 | 156.3 ± 0.2 |
| (1,1,2) + (2,2,1) | 162.4 ± 1.9 | 162.0 ± 0.8 |
| (2,1,1) | 167.5 ± 2.0 | ... |
| (1,2,1) | 177.7 ± 2.1 | ... |
| (1,1,1) | 182.9 ± 2.1 | 184.0 ± 1.8 |

TABLE IX. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (e)].

| Mode | Calculated value in μs | Measured value in μs |
|-------------------|-----------------------------------|---------------------------------|
| (2,2,2) | 141.8 ± 1.4 | 139.3 ± 4.0 |
| (2,1,2) | 150.9 ± 1.5 | 150.3 ± 1.2 |
| (1,2,2) + (2,2,1) | 157.2 ± 1.6 | 156.6 ± 0.6 |
| (1,1,2) + (2,1,1) | 167.5 ± 1.7 | 167.3 ± 1.1 |
| (1,2,1) | 172.5 ± 1.7 | ... |
| (1,1,1) | 182.9 ± 1.8 | 182.6 ± 1.7 |

TABLE X. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (f)].

| Mode | Calculated value in μs | Measured value in μs |
|-------------------|-----------------------------------|---------------------------------|
| (2,2,2) | 141.8 ± 1.4 | 139.0 ± 4.0 |
| (2,2,1) | 150.9 ± 1.5 | 151.3 ± 1.5 |
| (2,1,2) + (1,2,2) | 157.2 ± 1.6 | 156.0 ± 0.6 |
| (1,2,1) + (2,1,1) | 167.5 ± 1.7 | 168.6 ± 0.5 |
| (1,1,2) | 172.5 ± 1.7 | ... |
| (1,1,1) | 182.9 ± 1.8 | 182.6 ± 1.7 |

TABLE XI. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (g)].

| Mode | Calculated value in μs | Measured value in μs |
|-------------------|-----------------------------------|---------------------------------|
| (2,2,2) | 141.8 ± 1.7 | 139.3 ± 4.0 |
| (2,2,1) | 146.9 ± 1.7 | 145.6 ± 0.8 |
| (1,2,2) | 157.2 ± 1.8 | 156.6 ± 1.5 |
| (2,1,2) + (1,2,1) | 162.4 ± 1.9 | 162.0 ± 0.6 |
| (2,1,1) | 167.5 ± 2.0 | 168.0 ± 0.9 |
| (1,1,2) | 177.7 ± 2.1 | ... |
| (1,1,1) | 182.9 ± 2.2 | 184.0 ± 1.8 |

midway between the two echoes associated to the modes M_i and M_{i+1} (Fig. 7).

In Table II the calculated and the measured values of the arrival times of the modes (1,2) and (2,1) are compared for different values of d . The measured values of t_{ij} are obtained from the wavetrains recorded in Fig. 8. The 45° apex angle wedge is a duraluminum sample ($L = 400.0 \pm 0.1$ mm). The very good agreement obtained between the two determinations makes valid the hypothesis on the process of mode conversion in forward scattering experiments.

When three ASF modes are generated (Φ small), the possibility of mode conversions are schematically illustrated in Fig. 9: on point G, the incident ASF mode $M1$ gives rise to the modes (1,1) and (1,2), and the incident ASF mode $M2$ to the three modes (2,2), (2,1), and (2,3). The incident mode $M3$ is “pseudo-transmitted” (or forward scattered) via the two modes (3,3) and (3,2). These different modes are identified on the experimental wavetrains reported in Fig. 10 which represent the temporal records obtained for different values of the distance d (100.0 ± 0.1 mm, 150.0 ± 0.1 mm, and 200.0 ± 0.1 mm, respectively). The sample is a duraluminum wedge of 400 mm in length and 35° in apex angle. As was done before, the calculated and the measured arrival times are compared in Table III for modes (1,2), (2,1), (2,3), and (3,2). The agreement is still very good.

Another possibility for verifying the processus of mode conversion is to increase the number of points G where the glue is stuck. Here it is considered the case where the glue is placed in two positions G_1 and G_2 on the tip of a 50° apex angle wedge of duraluminum such that the two first ASF modes are generated only. Eight different combinations were studied (see Table IV). The corresponding wavetrains are replotted in Fig. 11. For these experiments, the identification of the forward modes requires three index (i, j, k). By considering that a mode conversion may occur on G_1 and G_2 , each of the indexes gives the type of the mode which propagates in the three successive sections of the wedge, respectively (AG_1 , G_1G_2 , and G_2B).

The calculated and measured values of the arrival times are compared for each of the identified modes in Tables V to XII for the different configurations. A very good agreement is again noticed. In the case of records 11(a) and (h), as the distance $G_1G_2 = 0$, only two modes have to be considered. It is noticed that the additional echoes are not observed when only one ASF mode ($M1$) is generated (see Fig. 12).

In conclusion, these quantitative experimental results re-

TABLE XII. Comparison between the calculated and the experimental values of the arrival time of the different echoes [curve (h)].

| Mode | Calculated value in μs | Measured value in μs |
|-------|-----------------------------|---------------------------|
| (2,2) | 141.7 ± 1.0 | 139.0 ± 4.0 |
| (2,1) | 157.3 ± 1.0 | 155.9 ± 1.5 |
| (1,2) | 167.2 ± 1.1 | 167.8 ± 0.5 |
| (1,1) | 182.9 ± 1.2 | 183.3 ± 1.8 |

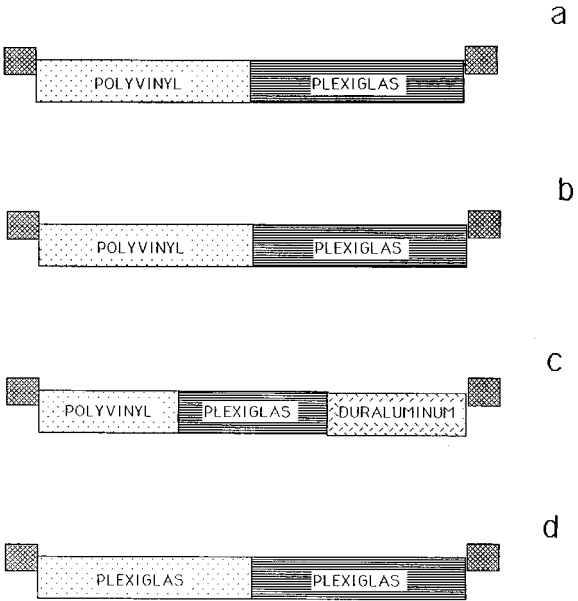
ported in this section and concerned with forward scattering confirm the mode conversion process as soon as a loading is exerted on the tip of the wedge in which the guided wave is propagating.

III. TRANSMISSION MEASUREMENTS

This section discusses with experimental data concerned by the transmission of edge waves propagating through wedges of different materials and apex angles. Samples 70.0 ± 0.1 mm in length were stuck together with a very thin layer of the fluid used for the coupling of the shear contact transducers. Several configurations were investigated; they are sketched in Fig. 13.

The wavetrains recorded in the four configurations are redrawn in Fig. 14. They are characterized by the existence of large signals, the origin of which has been successfully identified. The other echoes are not yet interpreted at the present time and necessitate additional experiments for a complete identification of the recorded echographic structure.

The wavetrain recorded with the first configuration [Fig. 13(a); $\Phi = 65^\circ$] is plotted in Fig. 14(a). The arrival time of the highest signal corresponds to the travel time of the fundamental ASF mode $M1$ which propagates in the two different media with the velocity of propagation associated to the crossed medium: The calculated value is equal to $142.9 \pm 1.0 \mu s$; the measured value is $141.4 \pm 2 \mu s$.



a/ Two different materials with the same apex angle ($\Phi = 65^\circ$).
b/ Two different materials with the same apex angle ($\Phi = 35^\circ$).
c/ Three different materials with the same apex angle ($\Phi = 65^\circ$).
d/ Same material with different apex angles ($\Phi = 65^\circ$ and 35°).

FIG. 13. Schematic diagrams describing the samples submitted to investigation.

In the second configuration [Fig. 13(b)], it is interesting to notice the presence of two large echoes [Fig. 14(b)]; they correspond to the transmission of modes $M1$ and $M2$. These two modes are transmitted directly through the interface and propagate in the different wedges ($\Phi = 35^\circ$) with the values rewritten in Table I. This is confirmed by the comparison between the calculated ($t_1 = 213.0 \pm 1.5 \mu s$ and $t_2 = 135.6 \pm 1.0 \mu s$) and measured values ($t'_1 = 210.0 \pm 1.8 \mu s$ and $t'_2 = 135.2 \pm 2.0 \mu s$) of the arrival times of modes $M1$ and

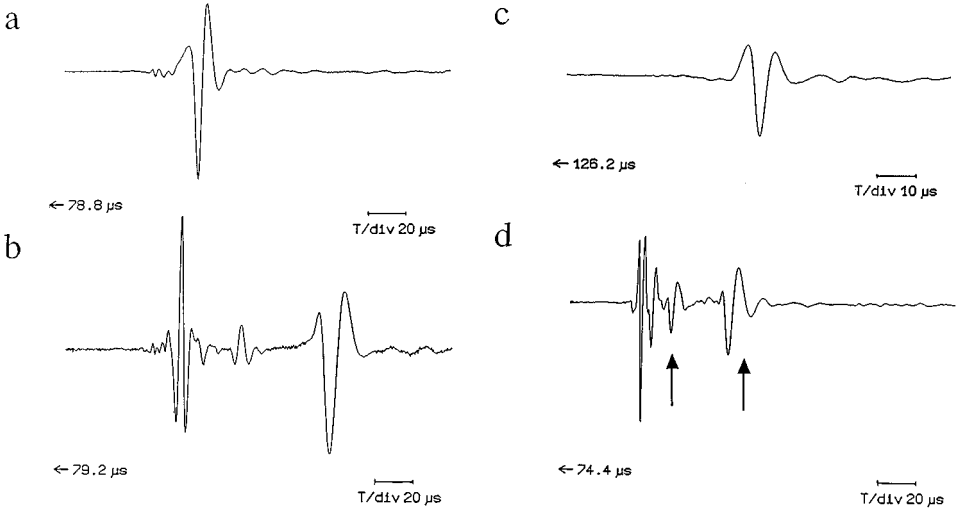


FIG. 14. Wavetrains associated to the different configurations described in Fig. 13.

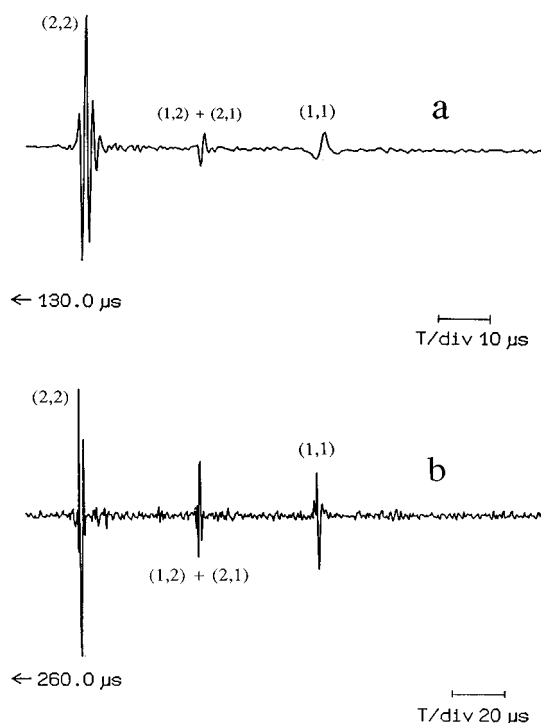


FIG. 15. Comparison between the wavetrains recorded in forward scattering geometry (a) and in reflection geometry (b). Notice the different scales of time between the two records.

$M2$, respectively. Once again, good agreement is obtained which confirms that the two modes are transmitted and propagate in each wedge with the velocity characteristic of the material for each mode.

The third experiment describes a generalization of the former measurements: Three samples of different materials (with a common apex angle $\Phi = 65^\circ$) are now coupled [13(c)]. It is easy to verify that the fundamental mode $M1$ propagates along the linear tip formed by the edges of the three wedges. The calculated and measured values of the travel time of mode $M1$ are 170.9 ± 0.8 and $170.8 \pm 1.0 \mu s$, respectively.

The last case [cf. Fig. 13(d)] is concerned with two samples of the same material (Plexiglas) but of two different wedge angles ($\Phi = 35^\circ$ and 65°). In this situation, the interpretation of the signals is more complicated [Fig. 14(c)]: The fundamental ASF mode only exists in the sample the wedge angle of which is 65° ; two ASF modes ($M1$ and $M2$) may exist in the other wedge ($\Phi = 35^\circ$). Then the question is, does a process of mode conversion exist at the interface as it was observed in the experiments previously described in this section?

If the two last echoes (marked with an arrow) are considered, it is noticed that their shapes are very similar (notice the slight dispersion). The experimental arrival times of these signals are 153.3 ± 1.8 and $122.4 \pm 1.3 \mu s$, respectively, and these values are close to the calculated values (155.6 ± 1.1 and $121.4 \pm 1.0 \mu s$) which correspond respectively to the propagation of the modes (1,1) and (1,2) through the two different wedges. So, it is concluded that a mode conversion exists at the interface between the two wedges. The existence

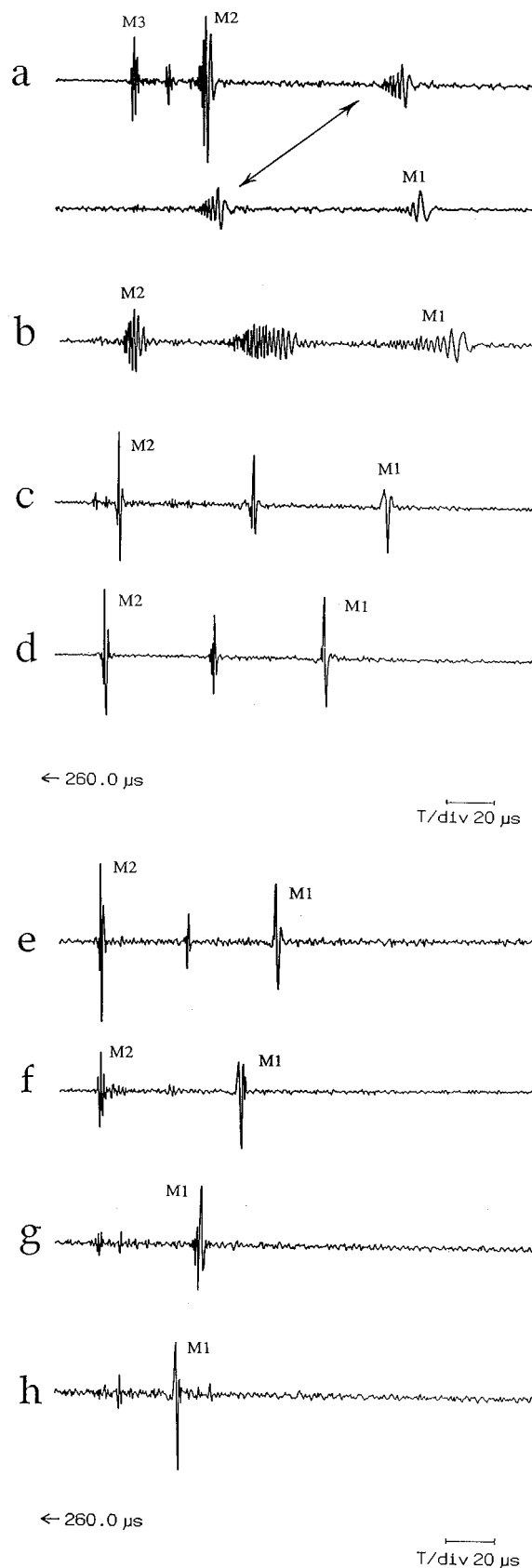


FIG. 16. Wavetrains recorded in reflection geometry on a duraluminum wedge ($L = 400$ mm) for different values of the apex angle: (a) $\Phi = 35^\circ$; (b) $\Phi = 40^\circ$; (c) $\Phi = 45^\circ$; (d) $\Phi = 50^\circ$; (e) $\Phi = 55^\circ$; (f) $\Phi = 60^\circ$; (g) $\Phi = 65^\circ$; (h) $\Phi = 70^\circ$.

of the two initial structures of this record may be justified by the fact that the coupling surface “ s ” is large¹² and it may result in different possibilities of conversion, including the conversion of ASF modes into a Rayleigh mode and conversely.¹⁰

IV. REFLECTION MEASUREMENTS

The geometry of the reflection measurements is schematically redrawn in Fig. 1(c). The wavetrains obtained with a 50° duraluminum wedge ($L=400\pm0.1$ mm) in forward scattering ($d=L/2$) and in reflection experiments are plotted for comparison in Fig. 15(a) and (b), respectively. These two temporal records reveal the same structure: The number of the echoes is the same and their positions are identical (notice that the time scale is multiplied by two for the record obtained in reflection geometry). In Fig. 15(a), the additional echo, which exists between the two main echoes (1,1) and (2,2), is due to a mode conversion (see Sec. III).

The comparison between the two records suggests that the same process of mode conversion exists when the incident mode is back scattered (or reflected) at the end of the free elastic wedge. In that case, the reflection is equivalent to a forward scattering experiment in which the glue should be positioned in the middle of the wedge. The amplitude of the signals may be different because the physical phenomena responsible of the mode conversion are not the same in both studies.

In Fig. 16 are replotted the wavetrains obtained by reflection at the free end face of duraluminum samples ($L=400\pm0.1$ mm) for different apex angles. It is interesting to notice that, as it was observed in forward scattering experiments, the signal identified resulting from a process of mode conversion disappears with mode $M2$ [Fig. 16(f)–(h)]. This result confirms that the process of mode conversion caused by the reflection at the free edge of the wedge does not exist when only one ASF mode is predicted.

V. CONCLUSION

The results reported in this paper present experimental information about the propagation and the scattering of edge

waves. In the case of forward scattering it is verified that a process of mode conversion occurs. This observation agrees with the theoretical predictions obtained by Krylov and Shanin for scattering by a notch. It is also confirmed that the reflection of a edge waves at the end of a free wedge gives rise to the generation of backscattered additional modes. Finally, the propagation and the conversion of low modes through different stuck samples is verified.

¹P. E. Lagasse, “Analysis of a dispersion free guide for elastic waves,” *Electron. Lett.* **8**, 372–373 (1972).

²S. L. Moss, A. A. Maradudin, and S. L. Cunningham, “Vibrational edge modes for wedges with arbitrary interior angles,” *Phys. Rev.* **8**, 2999–3008 (1973).

³V. V. Krylov, “Wedge acoustic waves: New theoretical and experimental results,” *Proceedings of the II International Symposium on Surface Waves in Solids and Layered Structures*, Varna, Bulgaria 1989 (World Scientific, Singapore, 1990), pp. 174–189.

⁴V. V. Krylov, “Propagation of wedge acoustic waves along wedges imbedded in water,” *Ultra. Sympo. Cat.* #94, CHO 793-796 (1994).

⁵J. R. Chamuel, “Arctic acoustic ultrasonic modeling studies,” Report JRC-10-90 (Sonoquest Advanced Ultrasonics Research), 1990, available from NTIC, Doc. No. A224 165/1/GAR (U.S. Dept. of Commerce).

⁶J. R. Chamuel, “Edge waves along immersed elastic elliptical wedge with range dependent apex angle,” *Ultra. Sympo. Cat.* #93, CHO 313-318 (1993).

⁷A. C. Hladky-Hennion, “Finite element analysis of the propagation of acoustic waves in waveguides,” *J. Sound Vib.* **194**, 119–136 (1996).

⁸X. Jia and M. de Billy, “Observation of the dispersion behavior of surface acoustic waves in a wedge waveguide by laser ultrasonics,” *Appl. Phys. Lett.* **61**, 2970–2972 (1992).

⁹X. Jia, D. Auribault, M. de Billy, and G. Quentin, “Laser generated flexural acoustic waves traveling along the tip of a wedge,” *Ultra. Sympo. IEEE Cat.* #93, 637-640 (1993).

¹⁰V. V. Krylov and V. Rajuzina, “Scattering of acoustic wedge modes,” *Sov. Phys. Acoust.* **34**, 546–547 (1988).

¹¹V. V. Krylov and A. V. Shanin, “Scattering of an acoustic wedge wave by a shallow notch,” *Sov. Phys. Acoust.* **39**, 155–158 (1993).

¹²M. de Billy, “Acoustic technique applied to the measurement of the free edge wave velocity,” *Ultrasonics* **34**, 611–619 (1996).

¹³W. Pajewski, M. Szalewski, and P. Kielczynski, “Ondes acoustiques sur l’arête d’un solide,” *Rev. Phys. Appl.* **22**, 113–118 (1987).

¹⁴J. R. Chamuel, “Flexural edge waves along free and immersed elastic waveguides,” in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 129–136.

A generalized internal source density method for the forward and backward projection of harmonic pressure fields from complex bodies

Peter R. Stepanishen

Department of Ocean Engineering, University of Rhode Island, Narragansett, Rhode Island 02882

(Received 6 September 1996; accepted for publication 27 February 1997)

A generalized internal source density (GISD) method is presented to address the forward and backward projection of harmonic pressure fields from complex three-dimensional bodies. The GISD approach is based on decomposing the field on a closed surface of revolution in the fluid into a summation of circumferential orders where the pressure field for each order is associated with an internal linear distribution of ring sources on the axis of revolution of the surface. Both Dirichlet and Neumann boundary value problems may be addressed using the approach in which the axial variation of each ring source distribution is formulated as the solution of a minimum mean square error problem which is subsequently solved using SVD methods. The resulting linear distributions of ring sources can be simply used to determine the entire pressure and velocity fields of a complex harmonically vibrating body. In particular, the pressure and/or normal velocity on a vibrating body can be readily obtained from the pressure field over a closed surface. Local intensity and radiated power can thus be simply obtained. Far field pressures are simply obtained from the axial Fourier transforms of the ring source distributions. Several examples are presented to illustrate the behaviour of the source distributions for spherical geometries. The analytical results obtained via the GISD method agree with those obtained via classical methods. © 1997 Acoustical Society of America. [S0001-4966(97)05106-0]

PACS numbers: 43.20.Fn [ANN]

INTRODUCTION

The forward and backward projection of a harmonic acoustic field from one region in space to another region is a basic problem in both air and underwater acoustics. Forward projection is defined here as the projection of a normal velocity or pressure field on a closed surface S_m surrounding an acoustic source as illustrated in Figure 1 to a more distant surface, whereas backward projection is defined here as the projection of the normal velocity or pressure field on S_m to a surface which is closer to the source and may in fact be the surface S of the acoustic source. A new mean squared error (MSE) approach is presented in the present paper to address the general three-dimensional forward and backward harmonic projection problem from complex three-dimensional bodies.

The present work is an extension of the internal source density approach previously reported by Stepanishen *et al.* to address acoustic harmonic radiation and scattering from two-dimensional bodies with a plane of symmetry¹⁻³ and to address axisymmetric projection problems for complex three-dimensional bodies of revolution.^{4,5} It is noted that the roots of the present approach can be attributed to von Karman⁶ who proposed the use of an internal monopole source density method in 1930 to address axisymmetric flow problems about axisymmetric bodies. Some recent related work was independently reported by Tomilina,⁷ Bobrovnikskii and Tomilina⁸ and Koopman *et al.*^{9,10} who also provided some additional history on the related method of superposition. Most recently, Ochmann¹¹ provided an excellent review of the related source simulation technique which replaces a ra-

diating surface or body via an equivalent set of simple sources within the body.

The general three-dimensional forward and backward projection problem of interest which is addressed here is first reduced to a series of two-dimensional problems via an expansion of the field in circumferential orders where the pressure field for each order is associated with an internal linear distribution of ring sources on the axis of a surface of revolution S_m . Both Dirichlet and Neumann boundary value problems on the closed surface of revolution S_m may be addressed using the generalized internal source density (GISD) approach in which the axial variation of each ring source distribution is formulated as the solution of a minimum mean square error problem which is subsequently solved using singular value decomposition (SVD) methods.

The resulting linear distributions of ring sources can be simply used to forward and/or backward project the field from the surface S_m and thus determine the entire pressure and velocity fields in the fluid surrounding a complex harmonically vibrating body. For the forward projection problem, far field pressures are simply obtained from the axial Fourier transforms of the ring source distributions. For the backward projection problem, the pressure and/or normal velocity on a vibrating body can be readily obtained from the external pressure field over a closed surface of revolution. Local intensity and radiated power can thus be simply obtained. The backward projection problem is thus important in various noise control and source identification studies and the GISD method presents an alternative to the use of various acoustic holographic techniques.¹²⁻²²

The general development of the GISD method is pre-

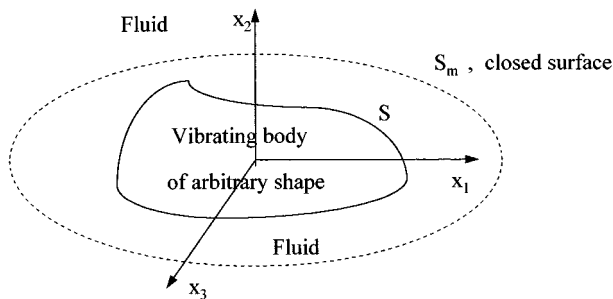


FIG. 1. An acoustic source or vibrating body in a fluid.

sented in the next section using an MSE method. Linear ring source distributions which are internal to S are introduced and are the building blocks or kernel of the approach. As a result of its importance, the evaluation of the kernel $K_n(\mathbf{x}, z_0)$, which is the pressure at \mathbf{x} due to the ring source at z_0 where n is the circumferential order of interest, is then addressed in the following section. An exact solution for the kernel is developed using a Green's function approach and an important small source approximation is introduced. Finally, several special examples are then addressed using the small source approximation in the last section. The results are shown to be in agreement with classical solutions to the problems.

I. GENERAL THEORY

Consider a complex shaped body with a surface S which is in contact with an external fluid as shown in Figure 1. The forward and backward projection of a harmonic acoustic field from one region in space to another region is now addressed for this problem where the harmonic normal velocity or pressure field on a closed surface of revolution S_m surrounding the body is presumed known from measurement or otherwise. Both the associated Neumann and the Dirichlet boundary value problems are addressed.

It is apparent that $p(\mathbf{x})$, the harmonic pressure field external to S , must satisfy the reduced wave equation in the external fluid volume V , i.e.,

$$\nabla^2 p(\mathbf{x}) + k^2 p(\mathbf{x}) = 0, \quad \mathbf{x} \in V, \quad (1)$$

where k is the acoustic wave number. A standard separation of variables in the spherical coordinates (R, θ, ϕ) shown in Figure 2 then leads to the following representation of $p(\mathbf{x})$

$$p(\mathbf{x}) = \sum_{n=-\infty}^{\infty} p_n(R, \theta) e^{jn\phi}, \quad \mathbf{x} \in V. \quad (2)$$

In addition, $p(\mathbf{x})$ must also satisfy the Sommerfeld radiation condition²³

$$\lim_{R \rightarrow \infty} R \left[\frac{\partial p}{\partial R} + jkp \right] = 0. \quad (3)$$

For the case of known Dirichlet boundary conditions on S_m it is obvious that $p(\mathbf{x})$ for $\mathbf{x} \in S_m$ can be decomposed into circumferential Fourier components as follows:

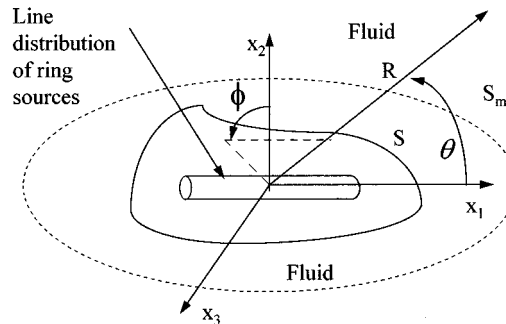


FIG. 2. A linear distribution of ring sources in a body.

$$p(\mathbf{x}(s)) = \sum_{n=-\infty}^{\infty} p_n(R(s), \theta(s)) e^{jn\phi}, \quad \mathbf{x}(s) \in S_m, \quad (4)$$

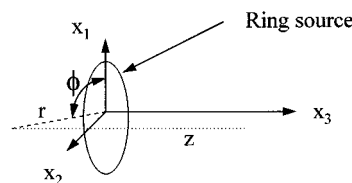
where s is a path parameter on S_m at $\phi=0$ and

$$p_n(R(s), \theta(s)) = \frac{\varepsilon_n}{2\pi} \int_{-\pi}^{\pi} p(\mathbf{x}(s)) e^{-jn\phi} d\phi, \quad (5)$$

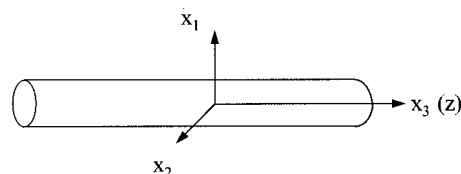
where ε_n is the Neumann symbol, i.e., $\varepsilon_0 = 1$ and $\varepsilon_n = 2$ for $n \geq 1$. For the Neumann boundary value problem, the specified normal velocity on S_m can be expressed in a similar manner.

In order to solve the Dirichlet boundary value problem represented by Eqs. (1)–(4), an internal linear distribution of ring sources along the axis of symmetry of S_m is introduced as shown in Figure 2 for each circumferential order of interest. For reasons which will become apparent, the source distributions are restricted to the region within S . The circumferential variation of each distribution of ring sources is matched to a circumferential mode of interest for the surface pressure in Eqs. (2). The axial variation of each ring source distribution is formulated as the solution of a minimum mean square error problem to satisfy the Dirichlet boundary condition in Eq. (4). Since the procedure to solve the analogous Neumann boundary value problem is identical, the details are presented here only for the Dirichlet problem.

Consider first a ring source of radius a in free space with a specified circumferential variation $e^{jn\phi}$ which is shown in Figure 3(A). The pressure field must satisfy the following inhomogeneous wave equation in cylindrical coordinates:



A. Ring source in free space



B. Linear distribution of ring sources

FIG. 3. A ring source and a linear distribution of ring sources.

$$\nabla^2 K_n(\mathbf{x}, z_0) + k^2 K_n(\mathbf{x}, z_0) = -\frac{\delta(r-a)\delta(z-z_0)}{r} e^{jn\phi}, \quad (6)$$

where $K_n(\mathbf{x}, z_0)$ is the pressure at \mathbf{x} due to the ring source at z_0 and n is the circumferential order of interest. The pressure field resulting from a line source distribution of ring sources shown in Figure 3(B) is thus

$$P_n(\mathbf{x}) = \int_{-L}^L K_n(\mathbf{x}, z_0) q_n(z_0) dz_0, \quad (7)$$

where $q_n(z)$ is the strength of the ring source at z and $q_n(z) = 0$ for $|z| > L$. It is readily apparent from the nature of the source distribution that the kernel $K_n(\mathbf{x}, z_0)$ and $P_n(\mathbf{x})$ are proportional to $e^{jn\phi}$ and can be expressed as:

$$P_n(\mathbf{x}) = P_n(\mathbf{x})|_{\phi=0} e^{jn\phi} \quad (8)$$

and

$$K_n(\mathbf{x}, z_0) = K_n(\mathbf{x}, z_0)|_{\phi=0} e^{jn\phi}. \quad (9)$$

It is obvious that $P_n(\mathbf{x})$ satisfies the radiation condition and the reduced wave equation external to the source region for all n . In order to solve the Dirichlet problem of interest $P_n(\mathbf{x})$ is now required to satisfy the Dirichlet boundary condition in Eq. (4) for all n via an appropriate determination of $q_n(z)$. A functional J_n is thus introduced to measure the error between the specified modal pressure $p_n(R(s), \theta(s))$ and that due to the internal line source distribution of ring sources $P_n(\mathbf{x})$ for $\mathbf{x} \in S_m$, i.e.,

$$J_n = \int_{S_m} \int |p_n(R(s), \theta(s)) e^{jn\phi} - P_n(\mathbf{x}(s))|^2 dS, \quad (10)$$

which, in light of Eq. (8), can be reduced to

$$J_n = 2\pi \int |p_n(R(s), \theta(s)) - P_n(\mathbf{x}(s))|_{\phi=0}|^2 r(s) ds, \quad (11)$$

where $r(s)$ is the radius of curvature in the circumferential direction. From Eqs. (7) and (11) it is thus apparent that $q_n(z)$ must satisfy the following integral equation of the first kind:

$$p_n(R(s), \theta(s)) = \int_{-L}^L K_n(\mathbf{x}(s), z_0)|_{\phi=0} q_n(z_0) dz_0. \quad (12)$$

The unknown distribution $q_n(z)$ may now be approximated using any of several methods²⁴ which then lead to the following generic representation for $q_n(z)$

$$q_n(z) = \sum_{i=1}^N w_i^{(n)}(z) q_i^{(n)}, \quad (13)$$

where the $w_i^{(n)}(z)$ may be a set of orthogonal functions over the length of the line source, e.g., Legendre functions. Alternatively, the $w_i^{(n)}(z)$ may be a suitable set of local basis or interpolation functions which are defined within a subinterval about z_i . As a limiting case of the latter class the $w_i^{(n)}(z)$ may be Dirac delta functions, i.e.,

$$w_i^{(n)}(z) = \delta(z - z_i). \quad (14)$$

It now follows from Eqs. (12) and (13) that $p_n(R(s), \theta(s))$ can be represented as

$$p_n(R(s), \theta(s)) = \sum_{i=1}^N K_i^{(n)}(s) q_i^{(n)}, \quad (15)$$

where the $K_i^{(n)}(s)$ are defined as follows:

$$K_i^{(n)}(s) = \int_{-L}^L K_n(\mathbf{x}(s), z_0)|_{\phi=0} w_i^{(n)}(z_0) dz_0. \quad (16)$$

A classical MSE solution of Eq. (15) then leads to the following normal equations,²⁴ i.e., a linear set of N algebraic equations for the N unknown source densities $q_i^{(n)}$

$$\sum_{i=1}^N a_{li}^{(n)} q_i^{(n)} = b_l^{(n)}, \quad l = 1, 2, \dots, N, \quad (17)$$

where

$$a_{li}^{(n)} = \int K_l^{(n)*}(s) K_i^{(n)}(s) r(s) ds \quad (18)$$

and

$$b_l^{(n)} = \int K_l^{(n)*}(s) p_n(R(s), \theta(s)) r(s) ds. \quad (19)$$

The radius of curvature $r(s)$ is included as a weighting function.

Although the solution of the normal equations can be readily obtained for those cases where ill conditioning of the coefficient matrix is not an issue, ill-conditioned matrices can readily occur due to the underlying integral equation of the first kind in Eq. (12). As a result of the often ill posed nature of the inversion for integral equations of the first kind, the singular value decomposition (SVD) method is more generally preferred^{24,25} to determine the N unknown source densities $q_i^{(n)}$. To cast the problem into a form suitable for the SVD method consider now the following set of algebraic equations which results from the use of Eq. (15) and an $M > N$ point collocation on the surface S_m :

$$\sum_{i=1}^N K_i^{(n)}(s_m) q_i^{(n)} = p_n(R(s_m), \theta(s_m)) \quad m = 1, 2, \dots, M. \quad (20)$$

Since Eq. (20) can be simply expressed in matrix form as follows:

$$[K] \mathbf{q}^{(n)} = \mathbf{p}_n, \quad (21)$$

where $[K]$ is an $M \times N$ matrix with $M > N$, SVD methods can then be used to represent $\mathbf{q}^{(n)}$ as follows:

$$\mathbf{q}^{(n)} = [K]^{-1} \mathbf{p}_n, \quad (22)$$

where $[K]^{-1}$ denotes the pseudo-inverse of $[K]$. For the overdetermined system of interest here where $M > N$, SVD produces a solution that is the best approximation in the least-squares sense,²⁴ i.e. the solution vector is determined so as to minimize $J^{(n)}$ where

$$J^{(n)} = \|[K] \mathbf{q}^{(n)} - \mathbf{p}_n\|^2. \quad (23)$$

The pressure field external to S_m in Figure 1 can now be represented as a summation of fields associated with line source distributions of ring sources within S , i.e.,

$$p(\mathbf{x}) = \sum_{n=-\infty}^{\infty} e^{jn\phi} \int_{-L}^L K_n(\mathbf{x}, z_0) |_{\phi=0} q_n(z_0) dz_0, \quad \mathbf{x} \in V. \quad (24)$$

In light of the approximation for $q_n(z)$ in Eq. (13), $p(\mathbf{x})$ can be then expressed as follows:

$$p(\mathbf{x}) = \sum_{n=-\infty}^{\infty} \sum_{i=1}^N T_{ni}(R, \theta) q_i^{(n)} e^{jn\phi}, \quad \mathbf{x} \in V, \quad (25)$$

where

$$T_{ni}(R, \theta) = \int_{-L}^L K_n(\mathbf{x}, z_0) |_{\phi=0} w_i^{(n)}(z_0) dz_0. \quad (26)$$

The associated velocity vector field $\mathbf{v}(\mathbf{x})$ can be simply evaluated using the linearized momentum equation and the local intensity vector can then be simply obtained, i.e.,

$$\mathbf{I}(\mathbf{x}) = p(\mathbf{x}) \mathbf{v}^*(\mathbf{x}). \quad (27)$$

If the pressure and velocity fields are evaluated over any closed surface S_w enclosing S , the radiated acoustic power W from the complex source can then be obtained as follows:

$$W = \frac{1}{2} \operatorname{Re} \left[\int_{S_w} \mathbf{I}(\mathbf{x}) \cdot d\mathbf{S} \right], \quad (28)$$

where \cdot denotes the dot product operation. Finally, for the case where the line sources are within the surface S of the source, the pressure and velocity fields on S_m can be backward projected, i.e., the fields can be analytically continued from S_m to the surface of the body S . The entire pressure and velocity fields external to the body can thus be simply obtained from the linear source density distributions.

II. THE KERNEL $K_n(\mathbf{x}, z)$

The kernel $K_n(\mathbf{x}, z_0)$, which is the pressure at \mathbf{x} due to the ring source at z_0 where n is the circumferential order of interest, was introduced in the preceding section and is an important building block for the generalized internal source density (GISD) method. It is readily apparent from the preceding section that the evaluation of the kernel $K_n(\mathbf{x}, z_0)$ is important in the numerical implementation of the generalized internal source density (GISD) method to address the forward and backward projection of harmonic pressure fields from complex three-dimensional bodies. The evaluation of the kernel is thus addressed in the present section.

Consider again the ring source in free space with a specified circumferential variation $e^{jn\phi}$ which is shown in Figure 3(A). The pressure field must satisfy the following inhomogeneous wave equation in cylindrical coordinates:

$$\nabla^2 K_n(\mathbf{x}, z_0) + k^2 K_n(\mathbf{x}, z_0) = - \frac{\delta(r-a) \delta(z-z_0)}{r} e^{jn\phi}, \quad (29)$$

where $K_n(\mathbf{x}, z_0)$ is the pressure at \mathbf{x} due to the ring source of radius a at z_0 and n is the circumferential order of interest.

After switching to spherical coordinates $\mathbf{x}=(R, \theta, \phi)$, it is apparent that $K_n(\mathbf{x}, z_0)$ can be expressed as:

$$\nabla^2 K_n(\mathbf{x}, z_0) + k^2 K_n(\mathbf{x}, z_0) = - \frac{\delta(R-a) \delta(\theta-\theta_0)}{a^2 \sin \theta_0} e^{jn\phi}, \quad (30)$$

where $z_0 = a \cos \theta_0$.

From a standard Green's function approach it is easily shown that

$$K_n(\mathbf{x}, z_0) = jk \rho_0 c_0 \int_V \int_V g(\mathbf{x}|\mathbf{x}_1) \frac{\delta(R_1-a) \delta(\theta_1-\theta_0)}{a^2 \sin \theta_0} e^{jn\phi_1} dV_1, \quad (31)$$

where the free space three-dimensional Green's function $g(\mathbf{x}|\mathbf{x}_1)$ can be expressed in spherical coordinates²⁶ as follows:

$$g(\mathbf{x}|\mathbf{x}_1) = - \frac{jk}{2\pi} \sum_{s=0}^{\infty} \sum_{l=0}^s \epsilon_l \frac{(s-l)!}{(s+l)!} (2s+1) \cos[l(\phi-\phi_1)] \times P_s^l(\cos \theta) P_s^l(\cos \theta_1) j_s(kR_1) h_s(kR), \quad R_1 \leq R, \quad (32)$$

where $j_s(\cdot)$ and $h_s(\cdot)$ are spherical Bessel and Hankel functions of the second kind respectively of order s , and $P_s^l(\cdot)$ are associated Legendre functions of the first kind. After using the sifting property²⁷ of the Dirac delta function to perform the integrations in Eq. (31) it then follows that

$$K_n(\mathbf{x}, z_0) = \frac{k^2 \rho_0 c_0}{4\pi} \sum_{s=0}^{\infty} \frac{(s-n)!}{(s+n)!} (2s+1) \times P_s^n(\cos \theta) j_s(ka) h_s(kR) e^{jn\phi} I_{ns}, \quad a \leq R, \quad (33)$$

where

$$I_{ns} = 2\pi P_s^n(\cos \theta_0). \quad (34)$$

Consider now the case of a ring source at $z_0=0$ or equivalently $\theta_0=\pi/2$. Since

$$P_s^n(0) = \frac{2^n}{\sqrt{\pi}} \cos \left[\frac{\pi}{2} (n+s) \right] \frac{\Gamma((n+s+1)/2)}{\Gamma((s-n)/2+1)} \quad (35)$$

for $n \geq 0$ where $\Gamma(\cdot)$ is the gamma function, then $K_n(\mathbf{x}, 0)$ can be expressed as follows:

$$K_n(\mathbf{x}, 0) = \frac{2^{n-1} k^2 \rho_0 c_0}{\sqrt{\pi}} \sum_{s=n}^{\infty} \frac{(s-n)!}{(s+n)!} (2s+1) \times P_s^n(\cos \theta) j_s(ka) h_s(kR) \times \cos \left[\frac{\pi}{2} (n+s) \right] \frac{\Gamma((n+s+1)/2)}{\Gamma((s-n)/2+1)} e^{jn\phi}. \quad (36)$$

As a result of the $\cos[(\pi/2)(n+s)]$ factor, it is noted that only those terms where $(n+s)$ is even are nonzero in Eq. (36).

For the case of $ka \leq 1$ it is first noted that

$$j_s(ka) \approx \frac{(ka)^s}{1 \cdot 3 \cdot 5 \cdots (2s+1)}, \quad ka \ll 1. \quad (37)$$

It then follows that $K_n(\mathbf{x}, 0)$ can be approximated for $n \geq 0$ by retaining only the $s = n$ term in Eq. (36):

$$K_n(\mathbf{x}, 0) \approx K_n \sin^n \theta e^{jn\phi} h_n(kR), \quad ka \ll 1, \quad (38)$$

where

$$K_n = (2n+1) \frac{(-1)^n k^2 \rho_0 c_0}{2\sqrt{\pi n!}} \frac{(ka)^n}{1 \cdot 3 \cdot 5 \cdots (2n+1)}. \quad (39)$$

It is noted that $K_n(\mathbf{x}, 0)$ is proportional to the sectoral surface harmonic $Y_n^n(\theta, \phi)$ which is defined as

$$Y_n^n(\theta, \phi) = e^{jn\phi} P_n^n(\cos \theta), \quad (40)$$

$$= e^{jn\phi} \sin^n \theta. \quad (41)$$

Repeating the above development for $n \leq 0$ it is easily shown that the following expression is valid for all n and $ka \ll 1$:

$$K_n(\mathbf{x}, 0) \approx K_{|n|} \sin^{|n|} \theta e^{jn\phi} h_{|n|}(kR), \quad ka \ll 1. \quad (42)$$

The kernel $K_n(\mathbf{x}, z_0)$ of interest is then simply obtained from Eq. (42) via a coordinate translation and may be expressed as follows for all n :

$$K_n(\mathbf{x}, z_0) \approx K_{|n|} \sin^{|n|} \theta(z_0) e^{jn\phi} h_{|n|}(kR(z_0)), \quad ka \ll 1, \quad (43)$$

where $\theta(z_0)$ and $R(z_0)$ are the spherical coordinates of a local coordinate system centered at z_0 on the axis of the internal source distribution.

III. SPECIAL CASES

Consider first the general projection problem where the field external to S_m in Figure 1 can be represented as a summation of fields associated with line source distributions of ring sources within S , i.e.,

$$p(\mathbf{x}) = \sum_{n=-\infty}^{\infty} e^{jn\phi} \int_{-L}^L K_n(\mathbf{x}, z_0) |_{\phi=0} q_n(z_0) dz_0, \quad \mathbf{x} \in V. \quad (44)$$

For the case of thin line source distributions of ring sources where $ka \ll 1$ it follows from Eq. (43) that $p(\mathbf{x})$ can be expressed as

$$p(\mathbf{x}) = \sum_{n=-\infty}^{\infty} K_{|n|} e^{jn\phi} \int_{-L}^L \sin^{|n|} \theta(z_0) h_{|n|}(kR(z_0)) \times q_n(z_0) dz_0, \quad \mathbf{x} \in V. \quad (45)$$

where

$$h_n(y) = j^{n+1} y^{-1} e^{-jy} \sum_{m=0}^n (n + \frac{1}{2}, m) (j2y)^{-m} \quad (46)$$

and

$$(n + \frac{1}{2}, m) = \frac{(n+m)!}{m! \Gamma(n-m+1)}. \quad (47)$$

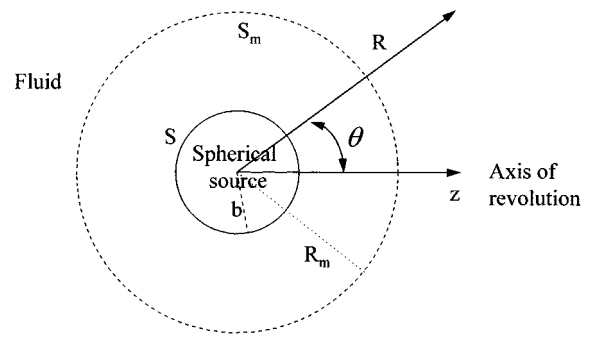


FIG. 4. A spherical source and spherical projection surface S_m .

In order to evaluate the far field of the line source distributions only the $m = 0$ term in Eq. (46) is used in Eq. (45). After the usual far field approximations in the resulting line integrals, it is easily shown that

$$p(\mathbf{x}) \sim \frac{e^{-jkR}}{kR} \sum_{n=-\infty}^{\infty} j^{|n|+1} e^{jn\phi} \sin^{|n|} \theta K_{|n|} \mathbf{Q}_n(k \cos \theta), \quad \mathbf{x} \in V, \quad (48)$$

where $\mathbf{Q}_n(k_z)$ is the Fourier transform of $q_n(z)$, i.e.,

$$\mathbf{Q}_n(k_z) = \int_{-L}^L e^{jk_z z_0} q_n(z_0) dz_0. \quad (49)$$

In contrast to the forward projection of the field on S_m to the far field, the backward projection of the field to S will, in general, require the use of Eq. (46) in Eq. (45).

As a test case for the GISD method to forward and backward project pressure fields, consider a vibrating sphere of radius b as shown Figure 4. The forward and backward projection of the pressure field is first illustrated here for the axisymmetric case where the normal or radial velocity of the sphere is specified as follows:

$$v(b, \theta, \phi) = V_n P_n(\cos \theta). \quad (50)$$

The associated pressure field is readily obtained via separation of variables which yields

$$p(R, \theta, \phi) = -j\rho_0 c_0 \frac{h_n(kR)}{h'_n(kb)} V_n P_n(\cos \theta). \quad (51)$$

More general axisymmetric fields are readily obtained via a summation over n .

Consider now the forward and backward projection of the pressure field from the spherical surface where $R = R_m$. It is apparent from Eq. (44) that the pressure field can be represented as

$$p(\mathbf{x}) = \int_{-L}^L K_0(\mathbf{x}, z_0) |_{\phi=0} q_0(z_0) dz_0, \quad \mathbf{x} \in V, \quad (52)$$

where the source distribution is to be determined from Eq. (12), i.e.,

$$-j\rho_0 c_0 \frac{h_n(kR_m)}{h'_n(kb)} V_n P_n(\cos \theta) = \int_{-L}^L K_0(\mathbf{x}(s), z_0) \big|_{\phi=0} q_0(z_0) dz_0, \quad (53)$$

where $L < b$ is required and $\mathbf{x}(s) = (R_m, \theta, \phi)$.

Several approximations are now introduced in order to analytically determine $q_0(z_0)$. The thin line approximation for the kernel in Eq. (43) first leads to the following integral equation for $q_0(z_0)$

$$-j\rho_0 c_0 \frac{h_n(kR_m)}{h'_n(kb)} V_n P_n(\cos \theta) = K_0 \int_{-L}^L h_0(kR(z_0)) q_0(z_0) dz_0. \quad (54)$$

If $q_0(z_0)$ is assumed to be concentrated about $z_0=0$, then $h_0(kR(z_0))$ can be approximated as follows:

$$h_0(kR(z_0)) \sim \frac{j e^{-jkR_m}}{kR_m} e^{jkz_0 \cos \theta}. \quad (55)$$

Subject to the condition that $kR_m \gg n$, it then follows that

$$\int_{-L}^L e^{jkz_0 \cos \theta} q_0(z_0) dz_0 = -j^{n+1} \rho_0 c_0 V_n \frac{1}{K_0 h'_n(kb)} P_n(\cos \theta). \quad (56)$$

In order to solve Eq. (56) for $q_0(z_0)$ it is first noted that $P_n(\cos \theta)$ is an n th order polynomial in $\cos \theta$, i.e.,

$$P_n(\cos \theta) = \sum_{s=0}^{E(n/2)} c_{ns} (\cos \theta)^{n-2s}, \quad (57)$$

where $E(n/2)$ is the integer part of $n/2$ and

$$c_{ns} = \frac{(-1)^s (2n-2s)!}{2^n s! (n-s)! (n-2s)!}. \quad (58)$$

After noting the following property of the p th derivative of the Dirac delta function,

$$\int_{-L}^L \eta(z_0) \delta^{(p)}(z_0) dz_0 = (-1)^p \frac{d^p \eta(z_0)}{dz_0} \bigg|_{z_0=0}, \quad (59)$$

it is apparent from Eq. (56) that

$$q_0(z_0) = -j^{n+1} \rho_0 c_0 V_n \frac{1}{K_0 h'_n(kb)} \sum_{s=0}^{E(n/2)} c_{ns} \frac{\delta^{(n-2s)}(z_0)}{(-jk)^{(n-2s)}}, \quad (60)$$

which has a concentrated region of support about $z_0=0$.

After substituting Eq. (60) into (52), the pressure field of interest can then be expressed as follows:

$$p(\mathbf{x}) = -j^{n+1} \rho_0 c_0 V_n \frac{1}{h'_n(kb)} \sum_{s=0}^{E(n/2)} \frac{c_{ns}}{(-jk)^{(n-2s)}} \times \int_{-L}^L h_0(kR(z_0)) \delta^{(n-2s)}(z_0) dz_0, \quad \mathbf{x} \in V. \quad (61)$$

The integrals in Eq. (61) are however readily evaluated via the use of Eq. (59) to obtain the following representation for $p(\mathbf{x})$

$$p(\mathbf{x}) = -j^{n+1} \rho_0 c_0 V_n \frac{1}{h'_n(kb)} \times \sum_{s=0}^{E(n/2)} \frac{c_{ns}}{(-jk)^{(n-2s)}} \frac{d^{n-2s} h_0(kR(z_0))}{dz_0^{n-2s}} \bigg|_{z_0=0}, \quad \mathbf{x} \in V. \quad (62)$$

Clearly, for $n=0$ Eq. (62) reduces to

$$p(\mathbf{x}) = -j\rho_0 c_0 V_0 \frac{h_0(kR)}{h'_0(kb)}, \quad (63)$$

which is the correct result in light of Eq. (51). For the $n=1$ case it is first noted that

$$\frac{dh_0(kR(z_0))}{dz_0} \bigg|_{z_0=0} = -kh_1(kR) \cos \theta. \quad (64)$$

Since $c_{10}=1$ from Eq. (58) it is apparent from Eq. (62) that

$$p(\mathbf{x}) = -j\rho_0 c_0 V_1 \frac{h_1(kR)}{h'_1(kb)} \cos \theta, \quad (65)$$

which is the correct result in light of Eq. (51).

As a final test case for the GISD method to forward and backward project pressure fields, consider now a harmonically translating sphere of radius b as shown in Figure 4 where the sphere is translating in the $\theta=\pi/2$ direction. The normal or radial velocity of the translating sphere is thus specified as follows:

$$v(b, \theta, \phi) = U_1 \sin \theta \cos \phi. \quad (66)$$

The associated pressure field is again readily obtained via separation of variables which yields

$$p(R, \theta, \phi) = -j\rho_0 c_0 \frac{h_1(kR)}{h'_1(kb)} U_1 \sin \theta \cos \phi. \quad (67)$$

Although the pressure field is noted to be axisymmetric about an axis in the $\theta=\pi/2$ direction, the field is clearly not axisymmetric about the $\theta=0$ direction which is the orientation of the line source distributions for the present problem.

Consider now the forward and backward projection of the pressure field from the spherical surface where $R=R_m$. For the case of thin line source distributions of ring sources where $ka \ll 1$ it follows from Eq. (43) that $p(\mathbf{x})$ can be expressed as

$$p(\mathbf{x}) = \sum_{\substack{n=-1 \\ n \neq 0}}^1 K_{|n|} e^{jn\phi} \int_{-L}^L \sin^{|n|} \theta(z_0) h_{|n|}(kR(z_0)) \\ \times q_n(z_0) dz_0, \quad \mathbf{x} \in V, \quad (68)$$

where $L < b$ is again required. Since

$$\cos \phi = \frac{e^{j\phi} + e^{-j\phi}}{2}, \quad (69)$$

matching the boundary condition at $R = R_m$ in Eq. (51) to the expression in Eq. (68) then leads to the following integral equation for $q_1(z_0)$

$$-j\rho_0 c_0 \frac{h_1(kR_m)}{2h_1'(kb)} U_1 \sin \theta \\ = K_{|1|} \int_{-L}^L \sin \theta(z_0) h_1(kR(z_0)) q_1(z_0) dz_0. \quad (70)$$

It is apparent from the procedure that $q_1(z_0) = q_{-1}(z_0)$.

Once again recalling the sifting property of the Dirac delta function, it is now obvious from Eq. (70) that $q_1(z_0)$ can be expressed as follows:

$$q_1(z_0) = \frac{-j\rho_0 c_0}{2K_{|1|} h_1'(kb)} U_1 \delta(z_0). \quad (71)$$

The forward and backward projection of the pressure field from the surface $R = R_m$ is then simply accomplished via the substitution of Eq. (71) into Eq. (68). It is easily shown, after performing the algebra, that

$$p(\mathbf{x}) = -j\rho_0 c_0 \frac{h_1(kR)}{h_1'(kb)} U_1 \sin \theta \cos \phi, \quad \mathbf{x} \in V, \quad (72)$$

which agrees with the result in Eq. (51).

IV. SUMMARY AND CONCLUSIONS

A new generalized internal source density (GISD) method is presented to address the forward and backward projection of harmonic pressure fields from complex three-dimensional bodies. The method is an extension of the mean squared error approach previously reported by Stepanishen^{4,5} to address axisymmetric projection problems for three-dimensional bodies of revolution. Briefly stated, the general three-dimensional forward and backward projection problem is reduced here to a series of two-dimensional problems via an expansion of the field in circumferential orders where the pressure field for each order is associated with an internal linear distribution of ring sources on the axis of a closed surface of revolution S_m . The axial variation of each ring source distribution is formulated as the solution of a minimum mean square error problem which leads to integral equations of the first kind which are subsequently solved using SVD methods.

The resulting linear distributions of ring sources can be simply used to forward and/or backward project the field from the surface S_m and thus determine the entire pressure and velocity fields in the fluid surrounding a complex harmonically vibrating body via simple line integrals. Far field

pressures are simply obtained from the axial Fourier transforms of the ring source distributions. Local velocity and intensity vector fields and radiated power can thus be simply obtained from the linear source distributions. In addition, the backprojected pressure, normal velocity and intensity on a vibrating body with a surface S can be readily obtained from the external pressure field over a closed surface of revolution S_m . Since the surface of revolution S_m can be of any shape, the GISD method is not restricted to coordinate surfaces for which the reduced wave equation is separable. The method thus offers significant advantages over existing planar and cylindrical acoustical holographic techniques of projecting acoustic fields, e.g., the surface S_m can be more closely matched to the surface S . It is noted that the backward projection problem is in general a more difficult problem due to the need of projecting evanescent components of the wave-field backward to the source surface S .

The kernel $K_n(\mathbf{x}, z_0)$, which is the pressure at \mathbf{x} due to the ring source at z_0 , where n is the circumferential order of interest, is an important building block for the generalized internal source density (GISD) method. As a result of its importance in the GISD method, the evaluation of the kernel was addressed in a separate section of the paper. Simple series expansions were developed using a Green's function approach; however, a particularly simple closed form expression was developed for the case where the ka of the ring source is small. The resulting expression is in fact a separable solution of the reduced wave equation in spherical coordinates and is important in the numerical implementation of the generalized internal source density (GISD) method to address the forward and backward projection of harmonic pressure fields from complex three-dimensional bodies. Although a set of constants K_n are introduced for the case where the ka of the ring source is small, the pressure field is independent of the constants since the source distributions are inversely proportional to K_n ; see Eq. (45).

Several special examples were presented to illustrate the nature of the axial behavior of the ring source distributions for both axisymmetric and non-axisymmetric problems from a spherical source. Closed form solutions to the problems were obtained for the source distributions which generally involve derivatives of Dirac delta functions centered at the origin of the spherical source. The associated pressure fields were shown to be equivalent to the more classical solutions which could be obtained via separation of variables. In light of the limited support region for the source distributions associated with the spherical geometry, the examples provided a verification of the GISD method for some nontrivial problems. In addition to the present examples, axisymmetric radiation and scattering problems involving both spherical and cylindrical bodies were addressed using the internal source density method in a recent paper²⁸ where the results were also shown to be in agreement with classical solutions.

Finally, it is noted that the GISD method leads to integral equations of the first kind which are often ill-conditioned. Numerical results using the GISD method for the case of general geometries were not presented here due to space limitations. Numerical results will however be presented in a following paper²⁸ for axisymmetric problems. For

these problems the line source distributions in the GISD method reduce to a single monopole line source distribution. A more detailed numerical investigation of the GISD method for nonaxisymmetric problems will be presented in a later paper to complement the present analytical study.

- ¹P. R. Stepanishen and S. Ramakrishna, "Acoustic radiation from cylinders with a plane of symmetry using internal multi-pole line-source distributions," *J. Acoust. Soc. Am.* **93**, 658–672 (1993).
- ²S. Ramakrishna and P. R. Stepanishen, "Acoustic scattering from cylinders with a plane of symmetry using internal multi-pole line-source distributions," *J. Acoust. Soc. Am.* **93**, 673–682 (1993).
- ³P. R. Stepanishen, "Acoustic two-dimensional radiation and scattering from cylinders using source density, SVD and Fourier methods," *J. Sound Vib.* (accepted for publication).
- ⁴P. R. Stepanishen and H. W. Chen, "Surface pressure and harmonic loading on shells of revolution using an internal source density method," *J. Acoust. Soc. Am.* **92**, 2248–2259 (1992).
- ⁵P. R. Stepanishen and H. W. Chen, "Acoustic harmonic radiation and scattering from shells of revolution using finite element and internal source density methods," *J. Acoust. Soc. Am.* **92**, 2248–2259 (1992).
- ⁶T. von Karman, "Calculation of pressure distribution on airships hulls," NACA TM 574 (1930).
- ⁷T. M. Tomilina, "Fast algorithm for sound field analysis based on fictitious sources method," *Int. Cong. Acoust.* **19**, 433 (1989).
- ⁸Y. I. Bobrovnikskii and T. M. Tomilina, "Application of ESM to a calculation of a finite elastic structure radiation," *Sov. Phys. Acoust.* **36**, 334–338 (1990).
- ⁹G. H. Koopman, L. Song, and J. B. Fahline, "A method for computing acoustic fields based on the superposition principle," *J. Acoust. Soc. Am.* **86**, 2433–2438 (1989).
- ¹⁰J. B. Fahline and G. H. Koopman, "A numerical solution for the general radiation problem based on the combined methods of superposition and singular-value decomposition," *J. Acoust. Soc. Am.* **90**, 2808–2819 (1991).
- ¹¹M. Ochmann, "The source simulation technique for acoustic radiation problems," *Acustica* **81**, 585–597 (1995).
- ¹²P. R. Stepanishen and K. Benjamin, "Forward and backward projection of acoustic fields using FFT methods," *J. Acoust. Soc. Am.* **71**, 803–812 (1982).
- ¹³E. G. Williams, J. D. Maynard, and E. Skudrzyk, "Sound source reconstruction using a microphone array," *J. Acoust. Soc. Am.* **68**, 340–344 (1980).
- ¹⁴E. G. Williams, H. D. Dardy, and R. G. Fink, "Near-field acoustical holography using an underwater automated scanner," *J. Acoust. Soc. Am.* **78**, 789–798 (1985).
- ¹⁵P. R. Stepanishen and H. W. Chen, "Forward and backward projection of acoustic fields from cylindrical vibrators," *J. Acoust. Soc. Am. Suppl.* **1** **74**, S24 (1983).
- ¹⁶J. D. Maynard, E. G. Williams, and Y. Lee, "Near-field acoustic holography. I. Theory of generalized holography and the development of NAH," *J. Acoust. Soc. Am.* **78**, 1395–1413 (1985).
- ¹⁷E. G. Williams, H. G. Dardy, and K. B. Washburn, "Generalized near-field acoustical holography for cylindrical geometry: Theory and experiment," *J. Acoust. Soc. Am.* **81**, 389–407 (1987).
- ¹⁸E. G. Williams, B. H. Houston, and J. A. Bucaro, "Broadband near-field acoustical holography for vibrating cylinders," *J. Acoust. Soc. Am.* **86**, 674–679 (1989).
- ¹⁹W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," *J. Acoust. Soc. Am.* **85**, 588–596 (1989).
- ²⁰G. V. Borgiotti, A. Sarkissian, E. G. Williams, and L. Scheutz, "Conformal generalized near-field acoustical holography for axisymmetric geometries," *J. Acoust. Soc. Am.* **88**, 199–209 (1990).
- ²¹G. V. Borgiotti and E. M. Rosen, "The determination of the far field of an acoustic radiator from sparse measurement samples in the near field," *J. Acoust. Soc. Am.* **92**, 807–818 (1992).
- ²²A. Sarkissian, "Near-field acoustical holography for an axisymmetric source," *J. Acoust. Soc. Am.* **88**, 961–966 (1990).
- ²³I. Stakgold, *Boundary value Methods in Mathematical Physics* (MacMillan, New York, 1966).
- ²⁴W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge Press, Cambridge, 1986).
- ²⁵R. H. Hanson, "A numerical method for solving Fredholm integral equations of the first kind using singular values," *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **8**, 616–622 (1971).
- ²⁶P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
- ²⁷A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1977).
- ²⁸P. R. Stepanishen, "Acoustic axisymmetric radiation and scattering from bodies of revolution using the internal source density and Fourier methods," *J. Acoust. Soc. Am.* (accepted for publication).

Ultrasonic anisotropic phase velocity determination with the Radon transformation

A. Kebaili^{a)} and D. R. Schmitt^{b)}

Institute of Geophysics, Meteorology and Space Physics, Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada

(Received 7 February 1996; accepted for publication 19 December 1996)

Coplanar arrays of small longitudinal ultrasonic transducers acquire a set of waveforms which are initially plotted as an amplitude image in the offset-time domain. This amplitude image is then mapped to the intercept time-horizontal slowness (τ - p) domain via a slant-stack Radon transformation. The shape of the locus of these arrivals in τ - p space directly provides information on the propagation angle-dependent phase velocities in an anisotropic medium. The method is applied to an isotropic acrylic and an anisotropic laminate with orthorhombic symmetry. The method allows determination of the phase velocities over an 80° range of propagation angles. The acrylic is isotropic with a phase velocity of 2780 m/s which differs little from that determined in a conventional pulse transmission. The phase velocities versus propagation angle were measured in three planes of symmetry and in one off-symmetry plane in the laminate. Anisotropies of up to 22% over an 80° range of propagation angles were found in the planes perpendicular to the layering. The plane containing the layering was only weakly anisotropic with velocities dependent on the weave of the fiber mats. © 1997 Acoustical Society of America. [S0001-4966(97)03205-0]

PACS numbers: 43.20.Gp, 43.58.Dj, 43.40.Le [JEG]

INTRODUCTION

Elastic properties provide important clues about the quality of both man-made and natural materials. Up to 21 independent elastic constants may be required to completely characterize any given material. These could, in principle, be obtained by an appropriate number of independent observations of the longitudinal and the shear wave phase velocities measured in the appropriate directions.¹ In practice, however, many more observations are required to invert velocities for elastic constants.²⁻⁵ Inversion methods usually lose generality by assuming that the measurements are made with respect to known symmetries within the material. Actual experiments are further complicated by difficulties in machining appropriate samples, by whether phase (plane) or group (ray) velocities are measured, and by travel-time picking errors. Resolving these experimental problems will improve determinations of a material's elastic properties.

Two types of velocities are defined in an anisotropic material: the ray (group) velocity which is the velocity of energy transport along the ray path, and the phase velocity defined as the velocity of a plane wave through the medium.^{6,7} The ray velocity is determined from the travel time between a point source and a point receiver at a known separation. This velocity is important in applications such as acoustic tomography where the ray paths through the object must be known to invert travel times for structure. In contrast, the elastic constants of a material are directly related to the phase velocity via Christoffel's equation^{1,6,7} and any attempt to determine the elastic tensor requires the knowledge of these phase velocities. Which of those velocities were

actually determined in a given experiment is not always clear.^{8,9}

Numerous ray velocity measurement methods have been proposed. For example, Markham¹⁰ describes a point-source, point-receiver technique which exploits Snell's law in determining velocity in an anisotropic material immersed in a fluid (water). Smith¹¹ determined the five elastic constants of carbon fiber composites, and Gieske and Allred¹² found the nine elastic constants of B-A1 composites using variations of this technique. A number of efforts have focused on the determination of the elastic moduli from the ray velocities.^{8,13,14}

Direct phase velocity measurements require that plane waves be propagated through the medium.⁹ In practice, this means that transducers must be large relative to the sample dimension through which the waves propagate. This limits the number of measurements obtained as the sample must be specially cut with numerous parallel faces to allow measurements in different directions. There have been a number of attempts to measure phase velocity. Rokhlin and Wang,¹⁵ and Mignogna¹⁶ demonstrated that phase velocities may be determined from Markham's¹⁰ transducer arrangement. The method relies on placing the detecting transducer at the position of maximum energy of the arriving pulse. In practice, the receiving transducer scans two orthogonal axes to find this position. This systemic scanning must be repeated for each incidence angle studied. Rokhlin and Wang¹⁷ eliminated this restriction by a double-through-transmission technique where the transmitter also receives the beam reflected back along the same path. However, a small nonparallelism of the sample can slightly change the refraction angle resulting in deviation of the acoustic path.¹⁷ Further, in highly attenuative samples the waveform shape differs from that of

^{a)}Currently at PanCanadian Petroleum Ltd., Calgary, Alberta, Canada.

^{b)}Electronic mail: doug@phys.ualberta.ca

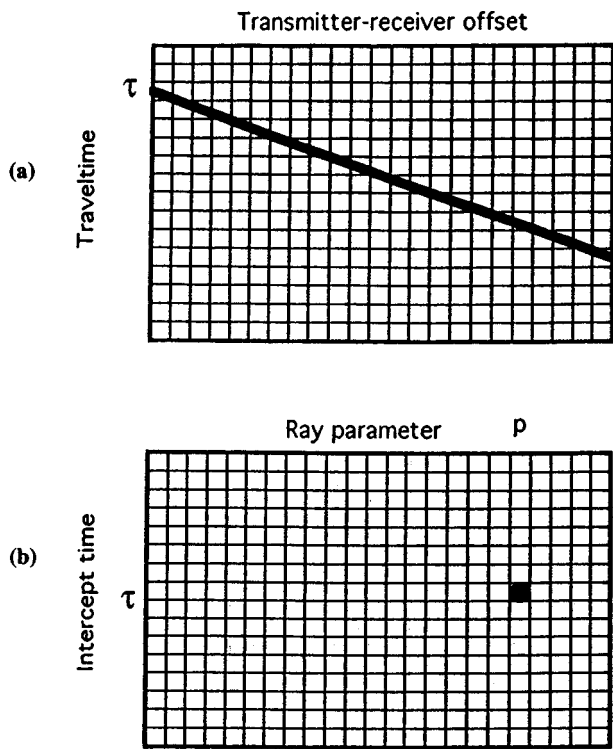


FIG. 1. Slant-stack procedure. The amplitudes in the offset-time space on a line of slope p and y -intercept τ are stacked (a) to give the amplitude in the τ - p space of x and y coordinates p and τ , respectively (b).

the reference making travel-time determination inaccurate and introducing error in the phase velocity.

Here, a special Radon transformation¹⁸ of a set of ultrasonic waveforms recorded with an array of point receivers placed on one surface of a sample is used to directly determine the phase velocity as a function of the phase angle of propagation. The transformation used here may essentially be considered as a plane-wave decomposition. The experimental geometry is greatly simplified in that the sample need only be machined as a rectangular prism. A differential analysis procedure is employed which further minimizes travel-time picking errors. The laboratory method described here is a modification of a technique developed to extract phase velocity anisotropy over depth intervals in the earth from wellbore seismic measurements.^{19,20} The method is described and tested on both an isotropic plastic and an orthorhombic fiber/epoxy matrix laminate.

I. THEORY

The Radon transform is a mapping of two-dimensional function $f(x, t)$ defined in the offset-time (x, t) coordinate system into a second (τ, p) system where τ and p are the t intercept and slope, respectively, of a line in the x - t plane. As defined by Robinson¹⁸ the Radon transform is equivalent to

$$F(\tau, p) = \int_{-\infty}^{\infty} f(x, \tau + px) dx, \quad (1)$$

which integrates the amplitudes $f(x, t)$ which lie along a line $y = \tau + px$; the value of this integration maps to the point

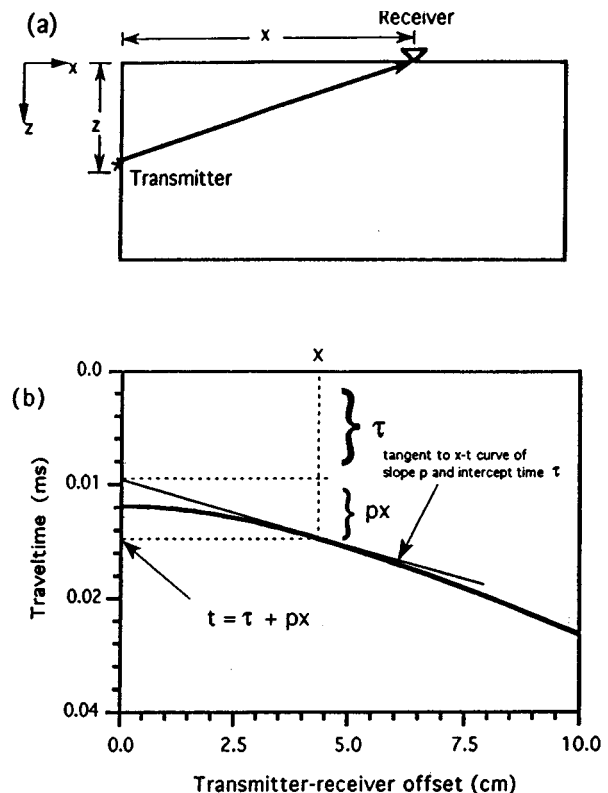


FIG. 2. (a) Ray path. (b) Through-transmission travel time versus transmitter-receiver offset for receivers aligned on the top surface and a transmitter on the side surface of a specimen.

(τ, p) in the τ - p domain (Fig. 1). In the discrete Radon transform²¹ the integral in Eq. (1) is replaced by summation

$$F(\tau_i, p_j) = \sum_{k=1}^n f(X_k, \tau_i + p_j X_k). \quad (2)$$

In the geophysical literature, this discrete transformation is often referred to as a slant stack as amplitudes are summed along lines of constant slope in the original x - t representation of the data.

Consider a prismatic sample as shown in cross section in Fig. 2(a). Disturbances generated by a point source on one surface of the specimen will reach in-line receivers placed on another surface along rays with angles of incidence which increase with offset. If the specimen is anisotropic, each disturbance travels at a different velocity dependent on the propagation angle. If the block is made of a typical isotropic plastic, the arrival times for a transmitter at a distance 2 cm from the surface to the receiving transducers along the top surface are described by the thick black hyperbolic line in Fig. 2(b). The slant stack essentially decomposes the locus of the directly arriving waveforms into the equivalent set of elementary plane waves according to their direction of propagation. This direction of propagation is represented by the horizontal slowness (ray parameter) $p = \sin \theta / v$ where θ is the angle of propagation. Alternatively, the ray parameter p is the instantaneous slope of the travel-time curve and as such the τ - p mapping is a plane-wave decomposition.²²

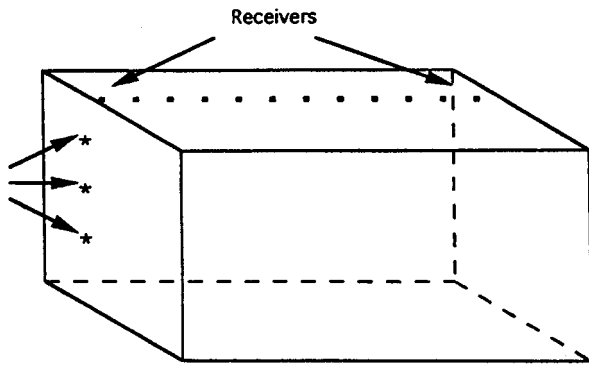


FIG. 3. Transmitter-receiver configuration for the experiments. The receivers are placed on the top surface of the specimen and the transmitters on the side surface in the vertical plane that contains the receivers.

Such a sorting of the received waveforms is particularly advantageous in anisotropy analysis as the velocity varies with θ which is implicit within p .

Equation (2) is employed to map ultrasonic waveforms observed by a series of in-line receiving transducers placed on the top sample's surface and at various distances from a transmitting transducer on the side (Fig. 3). The coordinates x and t are the distances x the receiving transducers are offset from the sample edge and the time t after the transmitting transducer has been activated, respectively. The two-dimensional function $f(x, t)$ is the amplitudes of the waveforms acquired by the receiving transducers at offset positions x with time t . The coordinates τ and p in Eq. (2) correspond to the intercept time and slope of a slant line in the $x-t$ domain, respectively. The two-dimensional function $F(\tau, p)$ is the mapped amplitudes of the transformed waveforms to the $\tau-p$ domain at intercept time τ and horizontal slowness p .

It is important to illustrate the differences between phase and ray velocities in describing why the $\tau-p$ mapping provides phase velocity at various incidence angles. Consider a source of elastic wave energy within an anisotropic medium activated at the origin (0,0) at time $t=0$. At a later time t the wavefront appears as shown in Fig. 4(a). An observer at point A' which is a distance x from the source position sees the arrival of the wave energy and calculates the ray (group)

velocity $V = x/t$. The wavefront W may alternatively be thought of as the envelope of the set of elementary plane-waves of all propagation directions simultaneously passing through the origin at the source activation time. The corresponding plane-wave-front tangent to W at A' falls along line $B-B'$ and is a minimum distance X from the origin. The phase velocity $v = X/t$. The angles of propagation within the material are Φ and θ for the ray and phase velocities, respectively.

The hypothetical phase slowness surface S for the medium in Fig. 4(b) contains point A at the same propagation angle θ . The length of the phase slowness vector s is inversely proportional to the speed of a plane wave propagating through the material at direction θ . The unit vector n is normal to S at A and is directed at the ray angle ϕ and is the direction of energy flow through the medium.

Consider now a ray propagating through an anisotropic medium. The travel time t of energy transport from a transmitter placed at the origin of a coordinate system to a receiver defined by a vector x of coordinates (x, y, z) can be written in terms of the phase slowness s [Fig. 4(b)] as

$$t = \mathbf{x} \cdot \mathbf{s}, \quad (3)$$

where \mathbf{s} is the phase slowness vector, or explicitly

$$t = x \frac{\partial t}{\partial x} + y \frac{\partial t}{\partial y} + z \frac{\partial t}{\partial z}. \quad (4)$$

Considering only the $x-z$ plane, Eq. (4) reduces to

$$t = xp + zq, \quad (5)$$

where $p = \partial t / \partial x$ and $q = \partial t / \partial z$ are the horizontal and vertical components of the phase slowness vector \mathbf{s} , respectively. Note that the travel time in Eq. (3) has been expressed in terms of the phase slownesses in Eq. (5). According to Snell's law $p = \sin \theta / v$ remains constant along the ray even if the medium is laterally heterogeneous. Note that in an anisotropic medium, however, the vertical slowness q is also dependent on the propagation direction through the material and consequently is also a function of p .

With this in mind, the travel time t for a given ray path may also be written (Fig. 2) as

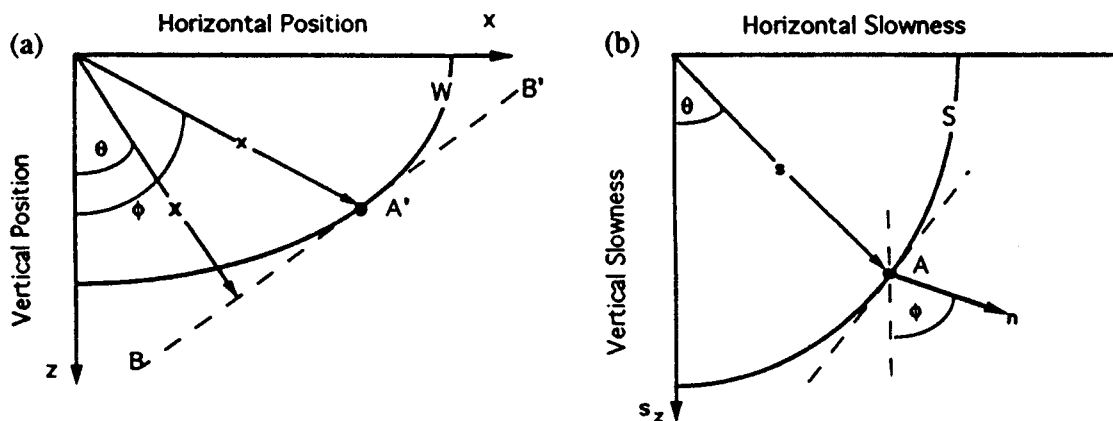


FIG. 4. (a) Hypothetical wave surface W in $x-z$ space at time t ; A' at the endpoint of spatial vector x at ray angle ϕ . (b) Hypothetical phase slowness surface S corresponding to W in one plane of an anisotropic material with respect to the vertical and horizontal.

$$t = \tau + xp. \quad (6)$$

By comparing Eqs. (5) and (6) the intercept time τ for each ray parameter p is equal to

$$\tau(p) = q(p)z. \quad (7)$$

The vertical component of the phase slowness q in Eq. (5) may then be obtained by dividing Eq. (7) by the distance of the transmitter from the top surface at the level of the receivers line.

The magnitude of the phase velocity v is the reciprocal of the slowness and is computed from the horizontal and vertical slowness components, p and q , respectively,

$$v = \frac{1}{|s|} = [p^2 + q^2(p)]^{-1/2} = \left[p^2 + \left(\frac{\tau(p)}{z} \right)^2 \right]^{-1/2}. \quad (8)$$

Equation (8) involves known quantities that allow the computation of the phase slowness.

As with picking of the first arriving energy in ultrasonic waveforms, the intercept time τ is not always easily found in a τ - p mapping. To minimize this problem, the difference between the τ - p curves for two transmitting transducers placed at different known distances from the receiving line on the top surface (Fig. 3) is used instead. If z_i is the distance from the top surface of the specimen for transmitter i , the intercept time τ_i for that transmitter is from Eq. (7),

$$\tau_i(p) = q_i z_i. \quad (9)$$

For two transmitters at distances z_1 and z_2 with $z_1 < z_2$, then for each ray parameter p , the intercept time τ_1 for the transmitter at distance z_1 can be subtracted from that of the transmitter at distance z_2 to yield an expression for the vertical slowness component q as a function of the transmitter spacing

$$q(p) = \frac{\tau_2(p) - \tau_1(p)}{z_2 - z_1}. \quad (10)$$

Equation (8) may then be adapted simply as

$$v(p) = \left[p^2 + \left(\frac{\tau_2(p) - \tau_1(p)}{z_2 - z_1} \right)^2 \right]^{-1/2}. \quad (11)$$

The advantage Eq. (11) presents over Eq. (8) is that a corresponding phase of the waveform (e.g., peak or trough) is selected for both τ - p mappings and the difference $\tau_2(p) - \tau_1(p)$ is then used in Eq. (11) to compute the phase velocities. Use of Eq. (11) assumes that the waveforms are not substantially altered by attenuation.

It is worthwhile to note that Eq. (11) provides a continuous measure of the velocity with the horizontal slowness p . This continuity of measurement allows ready conversion between phase and group velocities which requires knowledge of the derivative of phase velocity with angle of incidence. In planes of material symmetry only the ray angle Φ and ray speed V are related to the phase angle θ and phase speed v ,²³ by

$$V^2(\Phi) = v^2(\theta) + \left(\frac{dv}{d\theta} \right)^2, \quad (12)$$

$$\Phi(\theta) = \theta + \tan^{-1} \left(\frac{1}{v} \frac{dv}{d\theta} \right). \quad (13)$$

Note that this technique may be used, however, to determine the phase speeds in any plane containing the transmitters and receivers.

II. EXPERIMENTAL CONFIGURATION

The analysis described above is tested on two materials. The first is an isotropic acrylic which was annealed after machining to relax any residual stresses. Velocities were also measured in the sample by the pulse transmission method using 2.54-cm-diam, 0.7-MHz piezoelectric ceramics along the three orthogonal directions (x , y , and z). The observed velocities of 2826, 2780, and 2795 m/s, respectively, confirm that the sample is isotropic to within experimental errors. The second is an orthotropic phenolic with three planes of symmetry.²⁴⁻²⁶ This industrial laminate is composed of canvas fibers arranged in two orthogonal orientations and bonded by a phenolic resin. Within the fiber mats, one set of fibers is woven perpendicular to a second set of straight fibers. By convention, here the x -axis coincides with the woven fibers, the y axis with the straight fibers, and the z axis with the planes perpendicular to the fibers.

Both samples were machined into right rectangular prisms. The plexiglass and phenolic samples have dimensions $9.7 \times 4.7 \times 13.6$ cm measured to an accuracy of 0.05 cm. The faces of the phenolic specimen were cut parallel to the three planes of symmetry.

Ultrasonic transducers (0.2 \times 0.2 cm) were cut from larger PZT5 piezoelectric ceramics (American Piezoceramic). The size of the transducers was kept as small as possible in order to reduce complications which arise from integration of the responses from larger aperture transmitters and receivers.²⁷⁻²⁸ The transducers were coupled to the samples with a silver paint and damped by backing with a tungsten-epoxy matrix. The same sensor configuration was used for both the isotropic and anisotropic samples. The sensor arrays consist of a line of equally spaced receiving transducers on one surface of the specimen and a number of transmitters on the adjacent perpendicular surface (Fig. 3). These coplanar receivers and transmitters all lie within the plane of interest.

For the isotropic acrylic sample only one plane was investigated. Twenty receivers were aligned on the top surface at a 0.5-cm spacing, and three transmitters were placed on the side surface at distances of 1, 2, and 3 cm below the line of receivers (Fig. 3).

Transducer arrays were placed along the three planes of symmetry ($x - z$, $y - z$, and $x - y$) of the orthorhombic phenolic and also one diagonal plane containing the z axis and oriented 43° from the $y - z$ plane. Four transmitters also 1, 2, 3, and 4 cm below the surface containing the receivers were used in each of these four arrays.

The transmitters were activated by a 300-V spike provided by a high voltage pulse generator receiver (JSR SYNERGETIC, model PR 35). The response of the receiving transducer was pre-amplified and band-pass filtered with a 0.3- to 10-MHz passband prior to digitization at an 8-ns sampling rate on a Tektronix TDS 520 digital oscilloscope. A

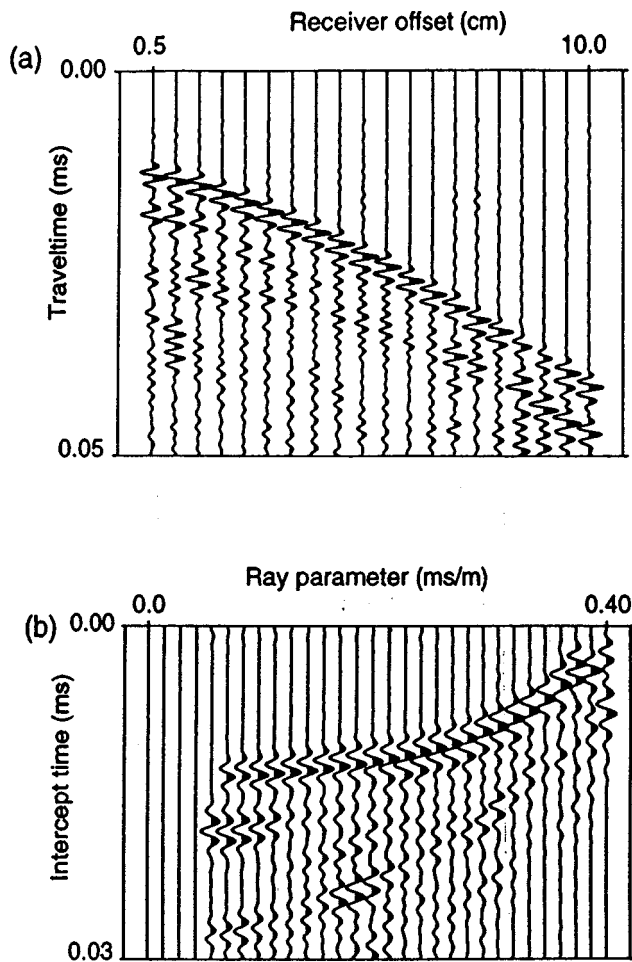


FIG. 5. (a) Waveform amplitudes with time observed for a transmitter on the side of the plexiglass specimen at 3 cm from the top surface where 20 receivers are placed along a line in the plane containing the transmitter. (b) $t-p$ mapping of the waveforms in (a). Positive values are shaded black to enhance viewing of coherent features.

final acquired waveform represents an average from at least 100 individual pulses. The final digitized waveforms were transferred to a Quadra 650 computer over the GPIB bus and to a SUN SPARC 5 workstation for analysis.

III. RESULTS AND DISCUSSION

A. Acrylic sample

Recorded waveforms which share the same transmitting transducer are first digitally band-passed filtered (300–850 KHz) and then plotted; an example of such an $x-t$ plot obtained on the plexiglass sample is shown in Fig. 5(a). The first arriving pulses follow the expected locus of a hyperbola in Fig. 5(a): These are the directly transmitted longitudinal waves through the sample. Other later arrivals include surface waves and the longitudinal waves reflected from the bottom surface. Using Eq. (2) the waveforms in Fig. 5(a) are mapped into the $\tau-p$ domain and plotted as shown in Fig. 5(b). Each data point in the $\tau-p$ domain is obtained by summing data samples in Fig. 5(a) along a straight line (slant-stack) tangent to the curve of slope p and y -intercept

τ . A tangency condition is satisfied by implementing a ratio filter, Moon *et al.*²⁹ This selective filter compares each data sample along the line of summation with its immediate neighbors at higher and lower p values. A datum is accepted into the summation only if the ratio of its amplitude with both neighboring samples is larger than a predetermined value. In our case this ratio is fixed at 0.5 for all $\tau-p$ mappings. Additionally, a velocity sensitive hyperbolic filter, Tatham³⁰ is introduced in the $\tau-p$ mapping to attenuate the edge effects with band-limited data; this method maps only those waveforms falling within a predetermined velocity range of $V_{\min}=2200$ m/s to $V_{\max}=3500$ m/s. Since ray parameter p and velocity V are related to each other by Snell's law, this restriction in velocity will allow each sample $f(x,t)$ in Eq. (2) to contribute to a certain number of slopes or p values in a range $p_{\min}-p_{\max}$. The initial results in the $\tau-p$ domain were finally bandpass filtered as above.

To determine the slowness curve of the specimen in the plane containing the transmitters and receivers, a coherent trace to trace feature (a trough or a peak) of the waveforms in the $\tau-p$ domain [Fig. 5(b)] is selected, and the intercept times $\tau_1(p)$ of this feature at each p is found. The intercept times $\tau_2(p)$ for the same feature on a corresponding $\tau-p$ mapping for a different transmitter are also found. The difference for each ray parameter p , $[\tau_2(p)-\tau_1(p)]$ is then calculated and divided by the spacing (z_2-z_1) of the two transmitters to yield the vertical component q of the phase slowness (Eq. 10).

In practice, the τ times are chosen for a number of different troughs and peaks in the $\tau-p$ waveforms in order to reduce random error. The resulting slowness $p-q$ curves can then be used to compute an average curve such as displayed in Fig. 6(a). The two symbols on the p and q axes of Fig. 6(a) are the directly measured slownesses for vertical and horizontal incidence angles. For this isotropic sample the data in Fig. 6(a) fall close along an expected circle of radius $|s|$; the solid line arc of circle in Fig. 6(a) represents the slowness curve calculated using this velocity. Further, Eq. (8) may be written as

$$q^2 = (1/|s|)^2 - p^2. \quad (14)$$

The graph q^2 vs p^2 is linear for the isotropic acrylic and has a y intercept equal to the reciprocal of the velocity of the specimen as shown in Fig. 6(b). Linear regression of the data in Fig. 6(b) yields a velocity of 2776 m/s which compares favorably with the $2830 \text{ m/s} \pm 23 \text{ m/s}$ measured independently by pulse transmission along the axes: This discrepancy is less than 1%.

The velocities computed using Eq. (11) are sensitive to the receiver spacing and the errors on the intercept time τ picks. The errors in recovered velocities are proportional to the reciprocal of the receiver spacing squared, and to the relative error on the intercept time difference as well. The larger the receiver spacing, the more accurate the recovered velocities. However, as the receiver spacing increases, the maximum propagation angle available decreases. One needs to find a balance between the range over which velocities may be determined and the accuracy of their determination.

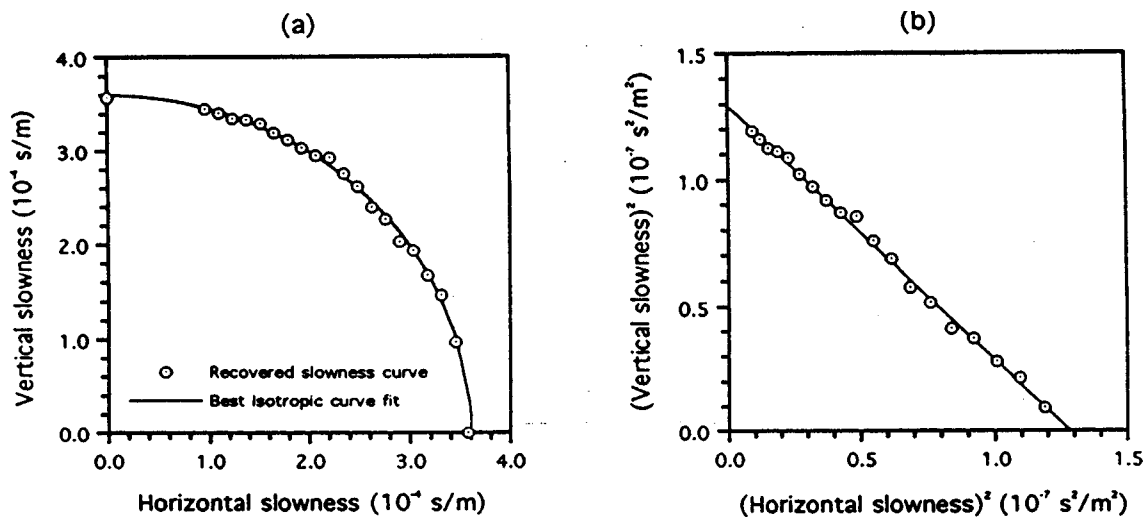


FIG. 6. Recovered (a) slowness surface and (b) velocity of the isotropic plexiglass sample.

B. Phenolic sample

The phenolic specimen is anisotropic with three planes of symmetry as described by Cheadle *et al.*²⁵ and more recently by Vestrum⁹ who used the pulse-transmission method to measure the travel-times which were then inverted to obtain slowness surfaces. Longitudinal and transverse phase velocities were first measured using large aperture transducers in nine directions on an 18 face specimen. Ray velocities were recovered from a 23-cm-diam. sphere of the same phenolic with the pulse-transmission method using small aperture transducers.

Here, the anisotropy of the sample is investigated in each of the three planes of symmetry and in one nearly diagonal plane aligned perpendicular to the x - y plane (plane of fibers) and at an angle of 43° with respect to the y - z plane (see Fig. 7). The data were analyzed using the same sequence as described above for the isotropic acrylic sample. As a reminder, the velocities are recovered from intercept time τ differences which makes the method less sensitive to the consistent travel time pick errors. The methodology is outlined for only the (y - z) plane parallel to the straight fibers but the final results for all planes will be presented. Two series of waveforms and their τ - p mappings corresponding to transmitters at 2 and 4 cm are shown in Figs. 8 and 9, respectively. The null traces at small ray parameters

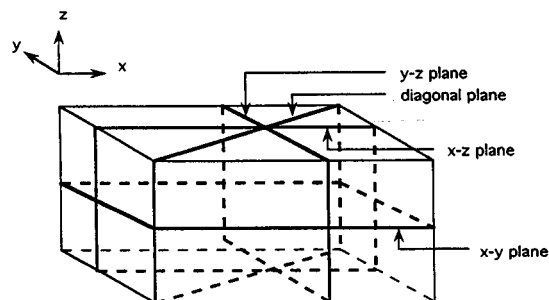


FIG. 7. Schematic of the phenolic sample showing the planes along which the longitudinal wave anisotropy was investigated.

p result from the hyperbolic filter. The intercept times τ for three different coherent lobes of these τ - p waveforms are selected and the intercept times τ for each of them are determined for both transmitters. Eq. (11) is finally used to compute the phase velocity for each ray parameter present in the data.

Figure 10 summarizes the phase velocity versus p functions obtained for the y - z plane from the plots in Figs. 8(b) and 9(b). The average phase velocity in the y - z plane increases with ray parameter from 2830 m/s along the z axis (perpendicular to the fibers) to 3300 m/s for the maximum ray parameter corresponding to a phase incidence angle θ of 80° : a velocity anisotropy of 17% over this range of incidence angles. An estimated uncertainty of 20 m/s is indicated by the variation in the velocities determined by picking of the different waveform features.

Having a near continuous measure of the phase velocity with ray parameter as in Fig. 10 is a considerable advantage in that the corresponding ray velocities may be determined using Eqs. (12) and (13). The ray parameters are easily converted to phase angles θ allowing $dv/d\theta$ in Eqs. (12) and (14) to be taken. The recovered phase velocity as a function of the phase angle θ was fit with a fourth order polynomial whose derivative is easily taken; this polynomial has no physical significance and was chosen for convenience only. Figure 11(a) shows that, as anticipated, the ray velocity is larger than the phase velocity for all ray angles between 15° and 80° .

The phase velocities recovered using the τ - p method are in good agreement with those directly measured by Vestrum⁹ with an average discrepancy of 1%. However, Fig. 11(a) shows that between incidence angles of 30° and 80° , Vestrum's ray velocities are systematically larger than those calculated here. This discrepancy may result from errors in distances and travel times or from slight differences from sample to sample but are consistent with Vestrum's estimated relative velocity error of 2%. The τ - p method used here to compute the velocities does not use the travel time through the sample but intercept time τ differences which

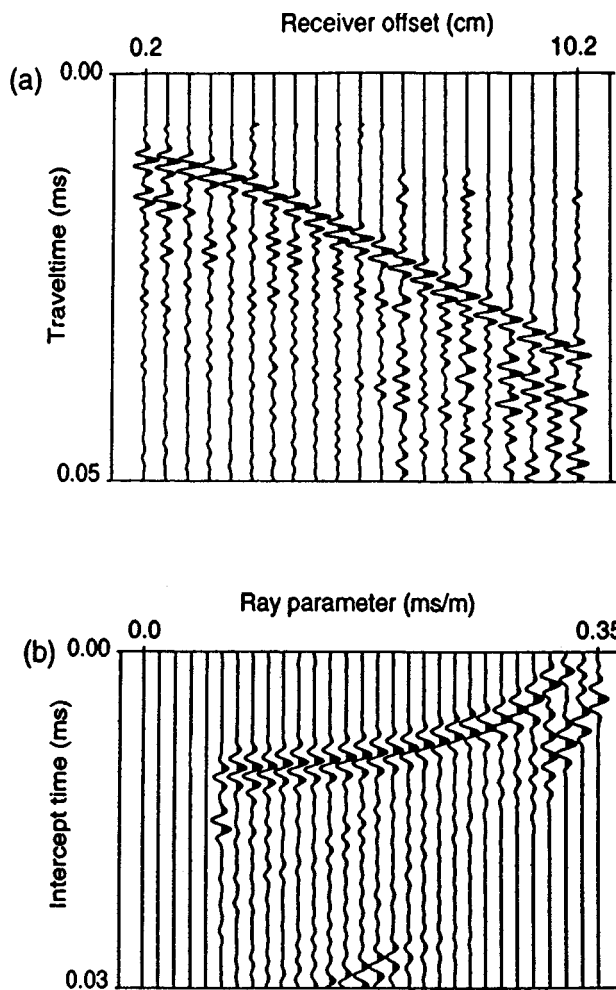


FIG. 8. (a) Waveforms observed for a transmitter on the side of the phenolic specimen at 2 cm from the top surface where 21 receivers are placed along a line in the $y-z$ plane that contains the transmitter. (b) $\tau-p$ mapping of the waveforms in (a). Areas under positive amplitudes are shaded black to enhance tracing of coherent features with p .

accounts for any systematic travel-time picking error.

The velocities along the axes are also measured using the transit-time method between a transmitter/receiver pair of transducers glued on opposite faces of the specimen. Opposing, 2.54-cm-diam, 1-MHz transducers were used here to measure the phase velocities in the present study, whereas small 1×1 mm, 1-MHz ceramics provided a measure of the ray velocities through a sphere.⁹ The measured phase and ray velocities along the axes are summarized in Table I.

Along the axes of symmetry the phase and ray velocities should be identical, but this is not the case here, especially for the y axis where the phase and ray velocities differ by as much as 3.5%. This discrepancy may result from travel-time measurement errors. One possible explanation for this discrepancy is that the larger transducers used to measure phase velocity provide a stronger signal which is less contaminated with noise than that for the smaller transducer used in the ray velocity measurement. Alternatively, the difference could arise from variations from sample to sample.

The phase and group velocities for the $(x-z)$ and

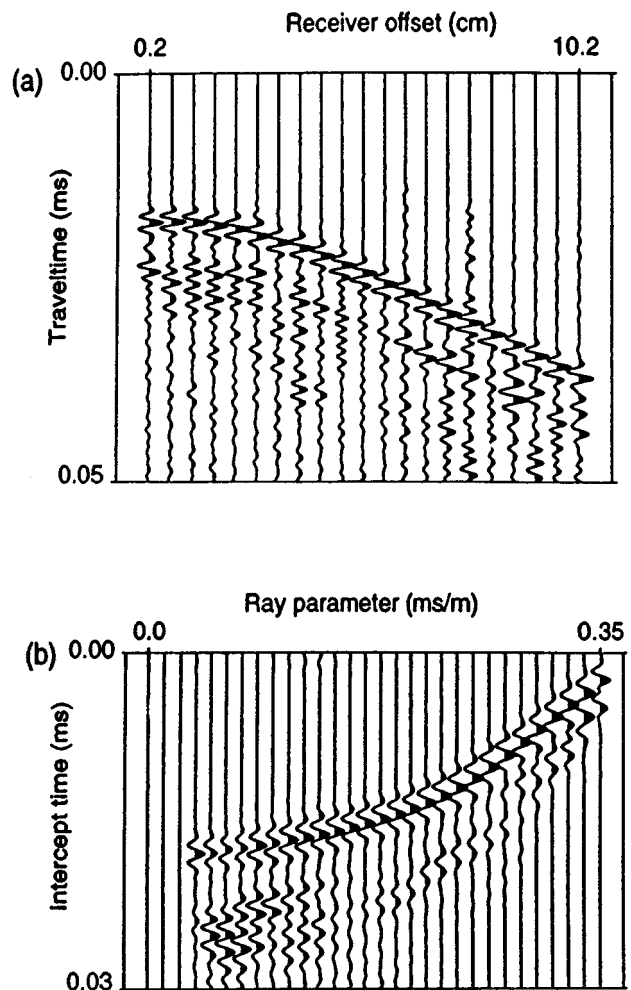


FIG. 9. (a) Waveforms observed for a transmitter on the side of the phenolic specimen at 4 cm from the top surface where 21 receivers are placed along a line in the $y-z$ plane that contains the transmitter. (b) $\tau-p$ mapping of the waveforms in (a).

$(x-y)$ planes of symmetry are shown in Fig. 11(b) and (c), respectively.

Figure 11(b) is very similar to that for the $y-z$ plane [Fig. 11(a)] and reveals a velocity anisotropy of 22%. The computed ray velocities are comparable to those measured

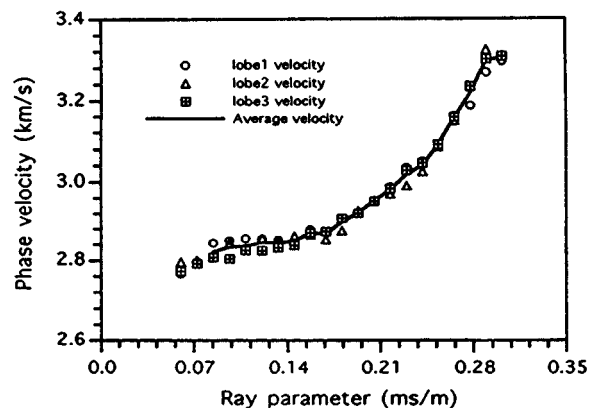


FIG. 10. Recovered phase velocities for the phenolic sample in the $y-z$ plane from the differences between $\tau-p$ curves in Figs. 8(b) and 9(b).

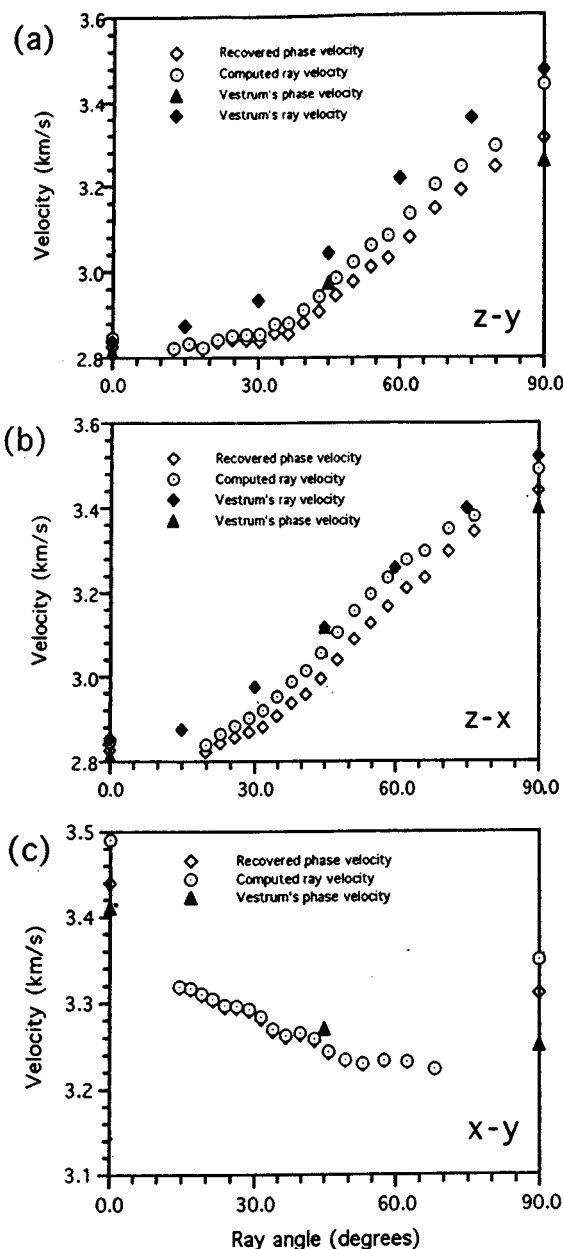


FIG. 11. Recovered phase (diamonds) and computed ray (open circles) velocities for the phenolic sample in the (a) $y-z$ plane, (b) $x-z$ plane, and (c) $x-y$ plane plotted as a function of the ray angle. The phase (filled triangles) and ray velocities as measured by Vestrum² are plotted for comparison. The angles are measured from the z axis (c) normal to the plane of the fibers for (a) and (b), and from the x -axis for the $x-y$ plane. In (c) the symbols for the recovered phase velocities are mostly covered by those for the computed ray velocities.

by Vestrum.⁹ However, Vestrum's phase velocity at 45° is 3% larger than our recovered phase velocity for the same angle using the $\tau-p$ method.

Figure 11(c) shows that the phase velocity in the $x-y$ plane decreases from 3320 m/s at 15° from the y axis to 3220 m/s for an incidence angle of 68° : a phase velocity change of only 3%. In this plane the computed ray and phase velocities differ little because of the weak anisotropy. The discrepancy of about 1.5% observed along the axes could again be due to the travel-time determination errors. The

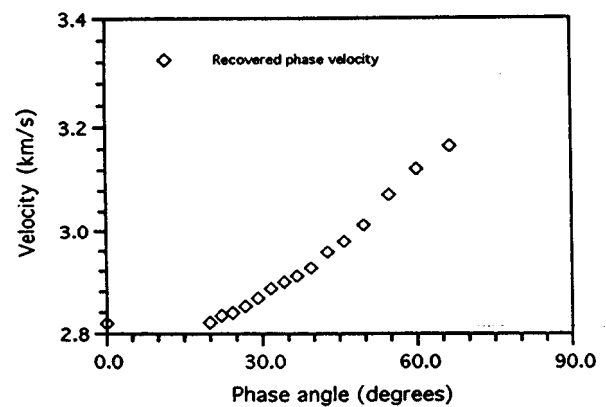


FIG. 12. Recovered phase velocities for the diagonal plane perpendicular to the $x-y$ plane plotted as a function of the phase angle.

slight increase in velocity from 70° to 90° was also observed by Vestrum.⁹

Unlike the other three planes, the diagonal plane is not a plane of symmetry. The phase velocities recovered from the $\tau-p$ curves relative to this plane show that the phase velocity increase from 2826 m/s in a direction perpendicular to the plane of the fibers to 3160 m/s for an incidence angle of 66° (Fig. 12). The ray velocities have not been computed for this off-symmetry plane as Eqs. (13) and (14) are valid in planes of symmetry only.^{6,7}

IV. CONCLUSION

The Radon transform is applied to ultrasonic waveforms plotted in the offset-time ($x-t$) space to the intercept time-horizontal slowness ($\tau-p$) domain. Since the horizontal slowness implicitly contains the angle of incidence, the $\tau-p$ domain is a natural space for velocity anisotropy determination. The intercept time τ is converted to the vertical slowness component when the transmitter spacing is known. The experimental transmitter-receiver configuration is particularly simple and easily implemented in a laboratory setting. Small transducers were coupled to the sample and backed with a tungsten-epoxy mixture for damping. Unlike the classic methods of acoustic phase velocity determination, which employ large transducers and are sometimes semidestructive if the sample must be cut in many different orientations, the $\tau-p$ method is nondestructive and provides a nearly continuous measure of the phase velocity with incidence angle.

On an isotropic plexiglass sample the $\tau-p$ method gave a velocity of 2780 m/s which differs from the through-transmission measurement of 2800 m/s by less than 1%. This technique was then applied to a phenolic sample of orthorhombic symmetry. The phase velocities were determined for four planes. The recovered phase velocities differ by less

TABLE I. Phase and ray velocities directly measured along the phenolic axes of symmetry.

| | y axis | x axis | z axis |
|----------------------|----------|----------|----------|
| Phase velocity (m/s) | 3310 | 3440 | 2826 |
| Ray velocity (m/s) | 3439 | 3490 | 2846 |

than 2% from those measured by Vestrum⁹ on a sample of the same material. Anisotropies of up to 22% were observed in the planes perpendicular to the layering with higher velocities observed parallel to the layers. This suggests the material is substantially less compliant in directions parallel to the layers. In contrast, a weak anisotropy of 3% in the plane of the fibers of the laminate sample was detected. Since the velocity curves are nearly continuous, the ray velocities in the planes of symmetry are easily computed.

Future work includes the development of small transducers capable of producing and detecting transverse particle motion. Such simultaneous longitudinal and shear velocities are necessary in determination of a material's elastic coefficients.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science and Engineering Research Council of Canada, the Alberta Oil Sands Technology and Research Authority, and the Petroleum Research Fund administered by the American Chemical Society. The laminate was supplied by Dr. J. Brown. Jay Haverstock assisted in the experiments.

- ¹B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973), Vol. 1.
- ²E. Schreiber, O. L. Anderson, and M. Soga, *Elastic Constants and their Measurements* (McGraw-Hill, New York, 1973).
- ³B. Hosten and B. Castagnede, "Optimization of the computation of elastic constants from the measurements of the ultrasonic velocities," *C. R. Acad. Sci. I Math.* **296**, 297–300 (1983).
- ⁴B. Castagnede, J. T. Jenkins, W. Sachse, and S. Baste, "Optimal determination of the elastic constants of composite materials from ultrasonic wave-speed measurements," *J. Appl. Phys.* **67**, 2753–2761 (1990).
- ⁵A. G. Every and W. Sachse, "Sensitivity of inversion algorithms for recovering elastic constants of anisotropic solids from longitudinal wavespeed data," *Ultrasonics* **30**, 43–48 (1992).
- ⁶M. J. P. Musgrave, *Crystal Acoustics* (Holden-Day, San Francisco, 1970).
- ⁷F. I. Fedorov, *Theory of Elastic Waves in Crystals* (Plenum, New York, 1968).
- ⁸R. W. Vestrum, "Group and phase-velocity inversions for the general anisotropic stiffness tensor," M.Sc. thesis, The University of Calgary, 1994.
- ⁹S. K. Sahay, R. A. Kline, and R. Mignogna, "Phase and group velocity considerations for dynamic modulus measurement in anisotropic media," *Ultrasonics* **30**, 373–383 (1992).
- ¹⁰M. F. Markham, "Measurements of the elastic constants of fibre composites by ultrasonics," *Composites* **1**, 145–149 (1970).
- ¹¹R. Smith, "Ultrasonic elastic constants of carbon fibers and their composites," *J. Appl. Phys.* **43**, 2555–2562 (1972).

- ¹²J. M. Gieske and R. E. Allred, "Elastic constants of B–Al composites by ultrasonic velocity measurements," *Exp. Mech.* **14**, 158–165 (1974).
- ¹³R. Kline and Z. Chen, "Ultrasonic technique for global anisotropic property measurement in composite materials," *Mater. Eval.* **46**, 986–993 (1988).
- ¹⁴A. G. Every and W. Sachse, "Determination of the elastic constants of anisotropic solids from acoustic wave group velocity measurements," *Phys. Rev. B* **42**, 8196–8205 (1990).
- ¹⁵S. I. Rokhlin and W. Wang, "Critical angle measurement of elastic constants in composite material," *J. Acoust. Soc. Am.* **86**, 1876–1882 (1989).
- ¹⁶R. B. Mignogna, "Ultrasonic determination of elastic constants from oblique angles of incidence in non-symmetry planes," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (New York, 1990), Vol. 9B, pp. 1565–1572.
- ¹⁷S. I. Rokhlin and W. Wang, "Double through-transmission bulk wave method for ultrasonic phase velocity measurement and determination of elastic constants of composite materials," *J. Acoust. Soc. Am.* **91**, 3303–3312 (1992).
- ¹⁸E. A. Robinson, "Spectral approach to geophysical inversion by Lorentz, Fourier and Radon transforms," *Proc. Inst. Electr. and Electron. Eng.* **70**, 1039–1053 (1982).
- ¹⁹D. R. Schmitt and A. Kebaili, "Velocity anisotropy estimation from slant stacks of wellbore seismics," *Can. J. Expl. Geophys.* **29**, 236–245 (1993).
- ²⁰A. Kebaili and D. R. Schmitt, "Velocity anisotropy observed in wellbore seismic arrivals: Combined effects of intrinsic properties and layering," *Geophysics* **61**, 12–20 (1997).
- ²¹C. H. Chapman, "Generalized Radon transforms and slant stacks," *Geophys. J. R. Astron. Soc.* **66**, 445–453 (1981).
- ²²P. L. Stoffa, P. Buhl, J. B. Diebold, and F. Wenzel, "Direct mapping of seismic data to the domain of intercept time and ray parameter—A plane-wave decomposition," *Geophysics* **46**, 197–209 (1981).
- ²³L. Thomsen, "Weak elastic anisotropy," *Geophysics* **51**, 1954–1966 (1986).
- ²⁴R. J. Brown, D. C. Lawton, and S. P. Cheadle, "Scaled physical modeling of anisotropic wave propagation: multioffset profiles over an orthorhombic medium," *Geophys. J. Int.* **107**, 693–702 (1991).
- ²⁵S. P. Cheadle, R. J. Brown, and D. C. Lawton, "Orthorhombic anisotropy: A physical seismic modeling study," *Geophysics* **56**, 1603–1613 (1991).
- ²⁶A. H. Nayfeh and D. E. Chimenti, "Ultrasonic plate waves in three-dimensional braided composites," *J. Acoust. Soc. Am.* **97**, 2056–2062 (1994).
- ²⁷B. Castagnede, K. Y. Kim, and W. Sachse, "Determination of the elastic constants of anisotropic materials using laser-generated ultrasonic signals," *J. Appl. Phys.* **70**, 150–157 (1991).
- ²⁸Y. C. Chu and S. I. Rokhlin, "Comparative analysis of through transmission ultrasonic bulk wave methods for phase velocity measurements in anisotropic measurements," *J. Acoust. Soc. Am.* **95**, 3204–3212 (1994).
- ²⁹W. Moon, A. Carswell, R. Tang, and C. Dillon, "Radon transform wave field separation for vertical seismic profiling data," *Geophysics* **51**, 940–947 (1986).
- ³⁰R. H. Tatham, "Multidimensional filtering of seismic data," in *Proc. IEEE* **72**, 1357–1369 (1984).

An efficient finite element scheme for solving the three-dimensional poroelasticity problem in acoustics

Raymond Panneton and Nouredine Atalla

G.A.U.S., Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada

(Received 18 October 1995; revised 6 November 1996; accepted 19 December 1996)

In this paper, the finite element method (FEM) is used to solve the three-dimensional poroelasticity problem in acoustics based on the isotropic Biot–Allard theory. A displacement finite element model is derived using the Lagrangian approach together with an analogy with solid elements. From this model, it is seen that the “damping” and “stiffness” matrices of the poroelastic media are complex and frequency dependent. This leads to cumbersome calculations for large finite element models and spectral analyses. To overcome this difficulty, an efficient algorithm is proposed. It is based on low-frequency approximations of the frequency-dependent dissipation mechanisms in poroelastic media. This efficient algorithm allows the poroelastic materials to be modeled with classical FEM codes. Also, the acoustic–poroelastic and the poroelastic–poroelastic coupling conditions are presented. The proposed model is compared to existing literature for both two-dimensional and three-dimensional problems. Excellent comparisons prove both the accuracy and effectiveness of the proposed model and its coupling with acoustic elements. Finally, to show the usefulness of the proposed model, the edge effect on the absorption coefficient of a multilayered poroelastic material is presented. Results show that both analytical models based on laterally infinite poroelastic layers, and standing wave tube measurements may be misleading at low frequencies.

© 1997 Acoustical Society of America. [S0001-4966(97)03305-5]

PACS numbers: 43.20.Gp, 43.50.Gf, 43.55.Dt [JEG]

INTRODUCTION

Sound absorbing materials are widely used in several noise control applications. Historically, these materials are known for their ability to dissipate acoustic waves propagating in their medium. They are often referred to as acoustic porous materials or absorbents. Extensive works have been done to characterize the acoustical behavior of porous materials. Two main classes can be drawn. The first one considers the porous material as an equivalent fluid with equivalent density and bulk modulus.^{1–4} This type of modeling applies to porous materials for which the frame motion can be neglected. Within these conditions, only one compression wave propagates in the air-saturated medium. Hence, Helmholtz’s equation is the governing equation in the porous medium.

The second class is more rigorous. It takes into account the elasticity of the skeleton. Hence, the porous medium bears three basic waves propagating simultaneously in the solid and fluid phases, namely, two compression and one shear waves. In this context, the word “poroelastic” is preferred to porous. The theoretical bases for the mechanical behavior of poroelastic materials have been mainly established by Biot (a review of the progression of Biot’s works on this subject can be found in Ref. 5). The so-called Biot theory⁶ describing the propagation of elastic waves in fluid-saturated porous media is of paramount importance in structural acoustics and vibration. Works by Allard *et al.*⁷ and Shiau⁸ were made to adapt specifically the Biot theory to acoustic problems. Following these works, the equations of dynamic equilibrium of poroelastic media were solved for simple one- and two-dimensional acoustic problems.^{7,9}

To study the effects of finite size absorbent treatment

and complex geometry, finite element formulations for sound absorbing materials have been developed. They used mainly equivalent fluid approaches.^{10,11} However, as previously mentioned, these approaches are limited to fibrouslike materials and/or situations in which the material is not directly bonded to the vibrating structure. Recently, to circumvent these limitations, finite element formulations based on the Biot theory have been proposed.^{12–15}

Kang and Bolton¹³ solved a two-dimensional coupled acoustic–poroelastic problem. The poroelastic medium was formulated in terms of the solid and pore fluid displacements (u - U). The starting point of this formulation was the differential equations of dynamic equilibrium given by Shiau⁸ for fluid-saturated porous media. Next, the discretized poroelastic equations were derived using the Galerkin’s method.¹⁶ Similarly, Johansen *et al.*¹⁴ presented a two-dimensional (u - W) finite element formulation to solve the coupled acoustic–poroelastic problem. The only difference with the Kang and Bolton’s formulation, lies in the use of the relative fluid–solid displacement (W) as the pore variable instead of the macroscopic fluid displacement. A similar (u - W) finite element formulation for the three-dimensional case was also proposed by Coyette and Wynendaele.¹⁵ Once again Galerkin’s method applied to the differential equations of dynamic equilibrium is used. However, the presented results were confined to one-dimensional and two-dimensional applications. Lately, a simplified approach using the solid displacement and the pore fluid pressure (u - P) as variables was proposed by Göransson.¹⁷ To enable the use of a (u - P) formulation, Göransson neglected the elastic coupling between the fluid and solid phases of the porous material. He solved

the one-dimensional wave propagation in porous media using a weighted residual formulation together with Galerkin's method. The solution was achieved in terms of the eigenfunctions satisfying the uncoupled wave equations. The main advantage of Göransson's formulation is the use of a reduced set of variables to describe the poroelastic material.

The aforementioned works mostly concentrate on the finite element formulation of the dynamic differential equations governing the motion of poroelastic media. No attention was devoted to the efficiency of the numerical implementation. This is understandable, since they discussed mainly simple one-dimensional and two-dimensional examples. However, for three-dimensional situations, the authors believe that the main drawback of the (u, U) formulation, lies in its heaviness since six degrees of freedom per node are required to describe the three displacements of the solid phase and the three displacements of the fluid phase. Consequently, it can lead to large frequency-dependent matrices. Without efficient numerical implementation this may hinder greatly the ability of such an approach to handle realistic configurations (e.g., airplane fuselage).

This paper addresses this problem. It describes an efficient three-dimensional $(u-U)$ finite element formulation for poroelastic media based on the Biot–Allard theory.¹⁸ Contrary to the previously mentioned works, the formulation is derived using the Lagrangian approach and uses an analogy with solid elements. This leads to the discretized equations of motion for poroelastic media and enables the use of classical FEM codes to handle poroelastic media; using solid elements. Also, and more importantly, it gives an insight into the matrix system to solve which in return leads to ways for boosting the efficiency of the algorithm. Accordingly, an efficient algorithm is presented, it is based on realistic low-frequency approximations of the frequency-dependent dissipation mechanisms in poroelastic media. Finally, the developed model is coupled with acoustic finite elements to tackle acoustic–poroelastic problems. Moreover, the coupling conditions between poroelastic media are given to handle multilayered poroelastic materials.

To validate the proposed model, several examples are discussed. Firstly, comparisons with exact analytical calculations of the surface impedance of laterally infinite single and multilayered poroelastic materials are given. They prove the accuracy of the finite element Biot–Allard model. Secondly, a comparison of the efficiency of a partial treatment of a rigid cubic cavity with analytical calculations is presented. It confirms both the efficiency of the acoustic–poroelastic coupling and the usefulness of the model to deal with practical three-dimensional situations. Finally, to show the usefulness of the proposed model, the edge effect on the absorption coefficient of a multilayered poroelastic material is presented. Results show that both analytical models based on laterally infinite poroelastic layers, and standing wave tube measurements may be misleading at low frequencies.

I. STATEMENT OF THE PROBLEM

The problem under consideration is depicted in Fig. 1. It consists of a three-dimensional finite domain made of poroelastic and acoustic subdomains. In Fig. 1, Ω_{pi} and Ω_a

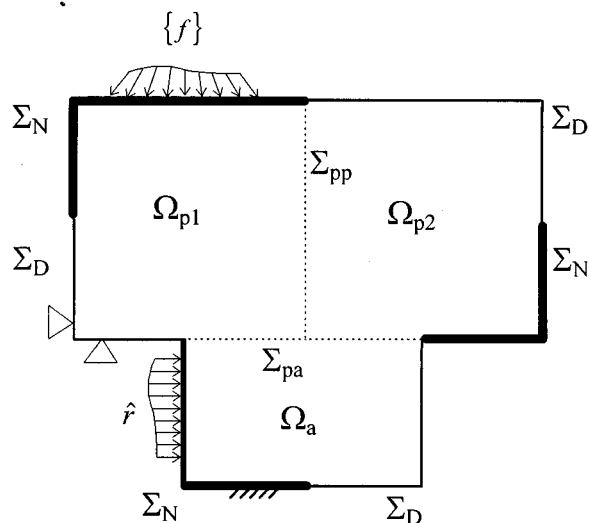


FIG. 1. Geometry of the problem.

represent the i th poroelastic subdomain and the acoustic subdomain, respectively. The surface of each subdomain is divided in three parts. One part is an interface with another medium, Σ_{pa} or Σ_{pp} , and the other two parts have Dirichlet and Neumann conditions, Σ_D and Σ_N , respectively. The field variable for the acoustic media is the acoustic pressure P . For the poroelastic media, the displacement vectors of the solid phase and the fluid phase, $\{u\}$ and $\{U\}$, are used as field variables. These displacements are average values in the sense of the Biot theory;⁶ i.e., the average volume displacements per unit cross-section area. Often, they are referred to as macroscopic displacements. The poroelastic subdomains are assumed homogeneous and isotropic. The application of the Biot theory⁶ to acoustics, as exposed by Allard,¹⁸ is used to model the poroelastic materials. In this study, this is referred to as the Biot–Allard model. Moreover, harmonic oscillations ($e^{j\omega t}$) is assumed. Finally, the poroelastic media are air-saturated and the air is initially at rest.

II. POROELASTIC FINITE ELEMENT MODEL

In the following, a Lagrangian approach together with finite element discretization is used to obtain the discretized equations of motion for a poroelastic medium. Also, for the sake of simplicity and for obtaining a generalized formulation which is compatible with commercial FEM codes, an analogy with solid elements is made; however, in the case of the poroelastic medium, six degrees of freedom per node are used. They account for three displacement components of the solid phase and three displacement components for the fluid phase.

A. Elasticity relations for poroelastic media

Defining $\{\sigma_s\} = (\sigma_{xx}^s \sigma_{yy}^s \sigma_{zz}^s \sigma_{xy}^s \sigma_{yz}^s \sigma_{zx}^s)^T$ and $\{\sigma_f\} = -hp\{m\}$ as the stress vectors of the solid and fluid phases, respectively, where h is the porosity of the poroelastic medium, p is the pore fluid pressure, and $\{m\} = (111000)^T$, the stress–strain linear relations for a poroelastic material are given by⁶

$$\{\sigma_s\} = [D_s]\{\epsilon_s\} + [D_{sf}]\{\epsilon_f\} \quad (1)$$

and

$$\{\sigma_f\} = [D_f]\{\epsilon_f\} + [D_{sf}]\{\epsilon_s\}, \quad (2)$$

in which $\{\epsilon_s\} = (\epsilon_{xx}^s, \epsilon_{yy}^s, \epsilon_{zz}^s, 2\epsilon_{xy}^s, 2\epsilon_{yz}^s, 2\epsilon_{xz}^s)^T$ and $\{\epsilon_f\} = (\epsilon_{xx}^f, \epsilon_{yy}^f, \epsilon_{zz}^f, 2\epsilon_{xy}^f, 2\epsilon_{yz}^f, 2\epsilon_{xz}^f)^T$ are the strain vectors for the solid and fluid phases, respectively, and $[D_s]$, $[D_f]$, and $[D_{sf}]$ are the matrices of the elastic coefficients related to the solid phase, the fluid phase, and the strain coupling between both phases, respectively. These matrices can be written as

$$[D_s] = \begin{bmatrix} A+2N & A & A & 0 & 0 & 0 \\ A & A+2N & A & 0 & 0 & 0 \\ A & A & A+2N & 0 & 0 & 0 \\ 0 & 0 & 0 & N & 0 & 0 \\ 0 & 0 & 0 & 0 & N & 0 \\ 0 & 0 & 0 & 0 & 0 & N \end{bmatrix}, \quad (3)$$

$$[D_{sf}] = Q\{m\}\{m\}^T, \quad (4)$$

$$[D_f] = R\{m\}\{m\}^T, \quad (5)$$

in which A and N are the Biot's coefficients corresponding to the Lamé's coefficients in the theory of elasticity. The expression of A is given in the Appendix. Here, a structural damping may be included by setting $[D_s] = [D_s]^*(1 + j\eta_s)$, where η_s is the loss factor of the frame, and $j = \sqrt{-1}$. Q is the Biot's coefficient representing the elastic coupling between the solid and fluid phases. It is equal to $(1 - h)^*K_f$, where K_f is the bulk modulus of the air in the pores. Similarly, R is the Biot's coefficient corresponding to the bulk modulus of the fluid phase of the aggregate and is equal to h^*K_f . Here, the bulk modulus of the air in the pores takes into account the dissipation due to the thermal exchanges between the fluid phase and the solid phase. Since the thermal exchanges depend upon the frequency, K_f is complex valued and frequency dependent. In this paper, the Champoux–Allard model⁴ is used to evaluate K_f . More details on the four elastic Biot's coefficients can be found in Ref. 18.

The strain–displacement relations for the solid phase and the fluid phase are, respectively,

$$\{\epsilon_s\} = [L]\{u\}, \quad (6)$$

$$\{\epsilon_f\} = [L]\{U\}, \quad (7)$$

where $[L]$ is the spatial derivative operator given in the Appendix, and $\{u\} = (u_x u_y u_z)^T$ and $\{U\} = (U_x U_y U_z)^T$.

B. Energy density functions

1. Kinetic energy

The kinetic energy density for a poroelastic medium is given by⁶

$$dT = \frac{1}{2}(\rho_{11}\{\dot{u}\}^T\{\dot{u}\} + 2\rho_{12}\{\dot{u}\}^T\{\dot{U}\} + \rho_{22}\{\dot{U}\}^T\{\dot{U}\}), \quad (8)$$

in which the dot over u and U denotes the time derivative d/dt , and ρ_{11} and ρ_{22} are the effective Biot's densities depending upon the geometry of the frame; they are related to the mass density of the frame and the fluid, and take into

account of the inertial interaction between both phases. The coefficient ρ_{12} quantifies this inertial interaction, it is related to the tortuosity of the frame. Expressions for these mass coefficients can be found in Ref. 18. These mass coefficients are frequency independent, as defined by Biot,⁶ and do not take into account the viscous effect. In this study, the viscous effect is taken into account by the complex-valued and frequency-dependent viscous damping coefficient derived from the works of Johnson *et al.*¹⁹ It will be shown later that the imaginary part of this damping coefficient can be interpreted as a modification of the effective Biot's densities.

2. Strain energy

The strain energy density for a poroelastic medium is given by

$$dU = \frac{1}{2}(\{\sigma_s\}^T\{\epsilon_s\} + \{\sigma_f\}^T\{\epsilon_f\}). \quad (9)$$

Substituting Eqs. (1)–(7) into Eq. (9) yields

$$dU = \frac{1}{2}([L]\{u\})^T[D_s][L]\{u\} + 2([L]\{u\})^T[D_{sf}][L]\{U\} + \{U\}^T([L]\{U\})^T[D_f][L]\{U\}. \quad (10)$$

The coupling between the strains of the frame and the fluid is represented by the middle term in Eq. (10).

3. Dissipation energy

The dissipation energy through viscous damping in the poroelastic material is introduced via the dissipation function,⁶

$$dD = \frac{1}{2}b(\omega)(\{\dot{u}\} - \{\dot{U}\})^T(\{\dot{u}\} - \{\dot{U}\}), \quad (11)$$

in which $b(\omega)$ is a complex and frequency-dependent viscous damping coefficient. It is related to the macroscopic flow resistivity and to the tortuosity of the porous material.⁸ It is noted that the dissipation is proportional to the relative motion between the solid phase and the fluid phase.

4. External work

Defining an external surface force $\{f\}$ on Σ_N , the work done per unit area by this surface force on the aggregate is given by²⁰

$$dW = \{u\}^T(\{f\} - h\{f_n\}) + h\{U\}^T\{f_n\}, \quad (12)$$

where $\{f_n\}$ is the normal force vector. In Eq. (12), the first two terms and the last term are the works related to the solid and fluid phases, respectively.

C. Finite element discretization

Discretizing the poroelastic domain into classical solid finite elements¹⁶ and interpolating the macroscopic displacement fields $\{u\}$ and $\{U\}$ in terms of the nodal displacements $\{\bar{u}\}$ and $\{\bar{U}\}$ yield, for the e th element,

$$\{u\}^e = [N_u]^e\{\bar{u}\}^e \quad \text{and} \quad \{U\}^e = [N_U]^e\{\bar{U}\}^e, \quad (13)$$

where $[N_u]^e$ and $[N_U]^e$ are the corresponding shape functions. Through Eq. (10), it is noted that these shape functions need to ensure C^0 continuity as for classical elastic solid problems.¹⁶

Substituting Eq. (13) into Eqs. (8) and (10)–(12) gives the discretized form of the energy densities for each element. Next, integrating over the e th element, summing the energy contribution of all the elements of the domain, and making use of Lagrange's equations with a dissipation function²¹ yield, for harmonic oscillations, the discretized equations of motion for the poroelastic medium,

$$\begin{aligned} & \left(-\omega^2 \begin{bmatrix} [M_{ss}] & [M_{sf}] \\ [M_{sf}] & [M_{ff}] \end{bmatrix} + j\omega \begin{bmatrix} [C_{ss}(\omega)] & -[C_{sf}(\omega)] \\ -[C_{sf}(\omega)] & [C_{ff}(\omega)] \end{bmatrix} \right. \\ & \left. + \begin{bmatrix} [K_{ss}] & [K_{sf}(\omega)] \\ [K_{sf}(\omega)] & [K_{ff}(\omega)] \end{bmatrix} \right) \begin{Bmatrix} \{\bar{u}\} \\ \{\bar{U}\} \end{Bmatrix} = \begin{Bmatrix} \{F_s\} \\ \{F_f\} \end{Bmatrix}, \quad (14) \end{aligned}$$

where

$$\begin{aligned} [M_{ss}] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [N_u]^e \rho_{11} [N_u]^e dV, \\ [M_{sf}] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [N_u]^e \rho_{12} [N_U]^e dV, \\ [M_{ff}] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [N_U]^e \rho_{22} [N_U]^e dV, \\ [K_{ss}] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [B_u]^e \rho_{11} [B_u]^e dV, \\ [K_{sf}(\omega)] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [B_u]^e \rho_{12} [D_{sf}(\omega)] [B_U]^e dV, \\ [K_{ff}(\omega)] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [B_U]^e \rho_{22} [D_{ff}(\omega)] [B_U]^e dV, \\ [C_{ss}(\omega)] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [N_u]^e b(\omega) [N_u]^e dV, \\ [C_{sf}(\omega)] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [N_u]^e b(\omega) [N_U]^e dV, \\ [C_{ff}(\omega)] &= \sum_{e \in \Omega_p} \int_{\Omega_p^e} [N_U]^e b(\omega) [N_U]^e dV, \\ \{F_s\} &= \sum_{e \in \Sigma_N} \int_{\Sigma_N^e} [N_u]^e (\{f\} - h\{f_n\}) dS, \\ \{F_f\} &= \sum_{e \in \Sigma_N} \int_{\Sigma_N^e} [N_U]^e h\{f_n\} dS, \end{aligned}$$

in which the strain–displacement matrices are

$$[B_u]^e = [L][N_u]^e \quad \text{and} \quad [B_U]^e = [L][N_U]^e.$$

In this matrix system, derived with the Lagrangian approach, the degrees of freedom related to the solid phase and the fluid phase are explicitly separated. Equation (14) can be written in the following condensed form:

$$(-\omega^2 [M_p] + j\omega [C_p(\omega)] + [K_p(\omega)]) \{\bar{W}_p\} = \{F_p\}, \quad (15)$$

where $[K_p]$, $[C_p]$, and $[M_p]$ are the equivalent “stiffness,” “damping,” and “mass” matrices of the poroelastic medium, $\{F_p\}$ is the nodal loading vector acting on the aggregate, and ω is the circular frequency. The unknowns are the nodal displacements $\{\bar{W}_p\}$. Similarly to the elasticity problem, the $(u-U)$ poroelasticity problem leads to a symmetrical system, Eq. (15). However, the “damping” and “stiffness” matrices of the poroelastic medium are complex valued and frequency dependent. This causes the related eigenvalue problem to be nonlinear in frequency, and prevents the use of a standard modal technique to solve Eq. (15). On the other hand, a direct solution technique can be used. However, for a spectral analysis, the building of the poroelastic matrix system, Eq. (15), will require cumbersome calculations since the “damping” and “stiffness” matrices have to be calculated at each frequency. To overcome this difficulty, astutenesses must be used and simplifications can be made as discussed further on.

III. ACOUSTIC FINITE ELEMENT MODEL

To model the acoustic domain, conventional acoustic finite elements are used. For the sake of conciseness, the global finite element formulation for the acoustic medium is directly given. For harmonic oscillations and an imposed normal displacement \hat{r} along the boundary Σ_N , it is given by²²

$$\left([H] - \frac{\omega^2}{c_0^2} [Q] \right) \{\bar{P}\} = \rho_0 \omega^2 \{R\}, \quad (16)$$

where

$$\begin{aligned} [Q] &= \sum_{e \in \Omega_a} \int_{\Omega_a^e} \{N_P\}^e \{N_P\}^{eT} dV, \\ [H] &= \sum_{e \in \Omega_a} \int_{\Omega_a^e} [B_P]^e [B_P]^e dV, \\ \{R\} &= \sum_{e \in \Sigma_N} \int_{\Sigma_N^e} \{N_P\}^e \hat{r} dS, \end{aligned}$$

in which

$$[B_P]^e = \nabla \cdot \{N_P\}^{eT}$$

and

$$P^e = \{N_P\}^e \{\bar{P}\}^e.$$

In Eq. (16), $[H]$, $[Q]$ are the “mass” and “stiffness” matrices of the fluid, P^e is the acoustic pressure on the e th element that is interpolated in terms of the element nodal pressure $\{\bar{P}\}^e$ using the shape function $\{N_P\}^e$, ρ_0 is the air density, and c_0 is the sound speed in air. A structural damping can be added in Eq. (16) by setting $[Q] = [Q](1 - j\eta_a)$, where η_a is the structural damping coefficient of the air. The right-hand side of Eq. (16) is the generalized excitation related to the prescribed volume displacement $\{R\}$.

IV. COUPLING CONDITIONS

The coupling conditions involved in the studied system are of the following types: acoustic–poroelastic and poroelastic–poroelastic. They are derived from Ref. 23.

A. Acoustic–poroelastic

The coupling conditions at the acoustic–poroelastic interface Σ_{pa} are

$$\begin{aligned}\{u_a\}^T \{n\} &= ((1-h)\{u\} + h\{U\})^T \{n\}, \\ [\mathcal{N}]^T \{\sigma_s\} &= -(1-h)P\{n\}, \\ [\mathcal{N}]^T \{\sigma_f\} &= -hP\{n\},\end{aligned}\quad (17)$$

where $\{u_a\}$ and P are the unknown acoustic displacement vector and pressure, respectively, $\{n\}$ denotes the unit outward normal to the acoustic medium along Σ_{pa} , and $[\mathcal{N}]$ is the matrix of the direction cosines of the unit outward normal along Σ_{pa} (see Appendix). The first equation ensures the continuity of the normal volume velocity (modified Euler's equation due to the porosity), while the second and third equations ensure the continuity of the normal stresses at the air-frame and the air-pore interfaces. For the poroelastic medium, the coupling with an acoustic medium may be viewed as a surface pressure at the interface Σ_{pa} . Consequently, it implies the addition of the following work term to Eq. (12):

$$dW_a = (1-h)P\{n\}^T \{u\} + hP\{n\}^T \{U\}. \quad (18)$$

Once the previous finite element procedure is reapplied to the poroelastic medium, taking into account Eq. (18) and using a proper interpolation for the acoustic pressure, a forcing term is added to the right-hand side of Eq. (14). It is given by

$$\begin{bmatrix} (1-h)[C_{sa}]\{\bar{P}\} \\ h[C_{fa}]\{\bar{P}\} \end{bmatrix}, \quad (19)$$

where the coupling matrices between air frame and air pore are, respectively,

$$[C_{sa}] = \sum_{e \in \Sigma_{pa}} \int_{\Sigma_{pa}^e} [N_u]^e \{n\}^e \{N_p\}^e dS,$$

$$[C_{fa}] = \sum_{e \in \Sigma_{pa}} \int_{\Sigma_{pa}^e} [N_U]^e \{n\}^e \{N_p\}^e dS.$$

Similarly, the harmonic motion imposed by the poroelastic medium on the acoustic medium along Σ_{pa} implies the addition of a new term to the right-hand side of Eq. (16). It is given by

$$\rho_0 \omega^2 ((1-h)[C_{sa}]^T \{\bar{u}\} + h[C_{fa}]^T \{\bar{U}\}). \quad (20)$$

B. Poroelastic–poroelastic

The coupling conditions at the poroelastic–poroelastic interface Σ_{pp} are

$$\begin{cases} \{u\}_1 = \{u\}_2 \\ h_1(\{U\}_1 - \{u\}_1)^T \{n\} = h_2(\{U\}_2 - \{u\}_2)^T \{n\}, \end{cases} \quad (21)$$

where subscripts 1 and 2 denote the two poroelastic media in contact. The first equation ensures continuity of the solid displacement vectors, while the second ensures the continuity of the normal component of the relative flow along the interface. These linear relations between the displacements are applied during the assembly process or using Lagrange multipliers.¹⁶ The continuity of the normal stresses is naturally taken into account in the energy formulation. Here, we only consider configurations where the pores of the two media are aligned at the interface. In this case, the flow through such an interface does not result in a pressure drop across the interface,²⁸ i.e., $P_1 = P_2$ along Σ_{pp} , where P_1 and P_2 are the pore pressures related to the poroelastic media Ω_{p1} and Ω_{p2} , respectively.

V. COUPLED SYSTEM

In the previous sections, the finite element formulations associated to each subdomain were developed in terms of the solid and fluid phase displacements for the poroelastic medium, and the acoustic pressure for the acoustic medium. Accordingly, for studying the coupled acoustic–poroelastic problem, the respective finite element formulations [Eqs. (15) and (16)] are coupled through Eqs. (19) and (20) and yield

$$\begin{bmatrix} -\omega^2[M_p] + j\omega[C_p(\omega)] + [K_p(\omega)] & -[C_{pa}] \\ -[C_{pa}]^T & \frac{1}{\rho_0 \omega^2} [H] - \frac{1}{\rho_0 c_0^2} [Q] \end{bmatrix} \begin{Bmatrix} \{\bar{W}_p\} \\ \{\bar{P}\} \end{Bmatrix} = \begin{Bmatrix} \{F_p\} \\ \{R\} \end{Bmatrix}, \quad (22)$$

where

$$[C_{pa}] = \begin{bmatrix} (1-h)[C_{sa}] \\ h[C_{fa}] \end{bmatrix}.$$

It is noted that the acoustic–poroelastic coupling is introduced using the coupling matrix $[C_{pa}]$. Equation (22) is solved using a direct solution method. It leads simultaneously to the nodal acoustic pressures in the acoustic me-

dium, and the nodal macroscopic displacements in the poroelastic medium. Note that modal analysis method with modal truncation may be used on the discretized equations related to the acoustic medium to reduce, partially, the size of the system. However, it cannot be used as such for the poroelastic medium. Actually, there is no easy way to extract the modes of the poroelastic medium by solving the related eigenvalue problem. Indeed, due to complex dissipation

mechanisms and the frequency-dependent nature of the structural and acoustical properties, the related eigenvalue problem for the poroelastic medium is complex, in the algebraic sense, and nonlinear in frequency. This problem prevents one from using a standard modal solution method with modal truncation to reduce the size of the poroelastic system.

VI. NUMERICAL IMPLEMENTATION

As noted in Eq. (15) or Eq. (22), the “damping” and “stiffness” matrices related to the poroelastic medium are complex valued and frequency dependent. For spectral analyses, this imposes the calculation of the global matrices $[C_p(\omega)]$ and $[K_p(\omega)]$ at each frequency step. Also, as previously mentioned, it prevents the use of a standard modal analysis method. Consequently, the chosen algorithm to solve Eq. (15) or Eq. (22) uses a direct solution method. This will lead to significant memory requirements and solution time for relatively large finite element models. Using the fact that the degrees of freedom of the solid and fluid phases are explicitly separated in the matrix system (15), and assuming the poroelastic medium to be homogeneous and isotropic, a more practical form for Eq. (15) writes

$$(-\omega^2[M_p] + j\omega b(\omega)[\hat{C}_p] + [K_1] + K_f(\omega)[K_2])\{\bar{W}_p\} = \{F_p\} \quad (23)$$

with

$$[K_1] = \begin{bmatrix} [K_{ss}] & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$[K_2] = \begin{bmatrix} 0 & (1-h)[\hat{K}_{sf}] \\ (1-h)[\hat{K}_{sf}] & h[\hat{K}_{ff}] \end{bmatrix}.$$

The hats over the matrices indicate that the damping coefficient and the bulk modulus of the air in the pores have been taken out from the volume integrals of Eq. (14). Contrary to Eq. (15), the “damping” and “stiffness” matrices in Eq. (23) do not need to be calculated at each frequency step. However, this does not allow for using a more efficient solution method but boosts the efficiency of the present algorithm since it reduces the number of calculations at each frequency step.

To further simplify the frequency dependence of Eq. (23), low-frequency approximations of the complex viscous damping coefficient and the bulk modulus of the air in the porous network may be used. The low-frequency approximation of the complex viscous damping coefficient is²⁴

$$b(\omega) = \beta + j\omega\alpha \quad (24)$$

with

$$\alpha = \frac{h^2\sigma}{2H}, \quad \beta = h^2\sigma, \quad (25)$$

and

$$H = \frac{\sigma^2\Lambda^2h^2}{4k_s^2\eta\rho_0}, \quad (26)$$

where σ , k_s , and h are the macroscopic flow resistivity, the tortuosity, and the porosity of the porous material, η is the viscosity of air, and H is the viscous characteristic frequency of the poroelastic medium which is related to the viscous characteristic dimension Λ introduced by Johnson *et al.*¹⁹ Equation (24) is obtained by keeping the first-order term in the series expansion of $b(\omega)$ in terms of the viscous dimensionless frequency ω/H . It is solely valid for $\omega/H \ll 1$. However, a convergence study, given in Ref. 24, shows that acceptable results are obtained even for values of ω/H near unity. The “exact” mathematical expression of $b(\omega)$ in the Biot–Allard model can be found in Ref. 18.

Similarly, the low-frequency approximation of the bulk modulus is²⁴

$$K_f(\omega) = P_a[1 + j\Gamma(\omega)] \quad (27)$$

with

$$\Gamma(\omega) = 2 \frac{(\gamma-1)}{\gamma} \cdot \frac{\omega}{H'} \quad (28)$$

and

$$H' = \frac{16\eta}{B^2\Lambda'^2\rho_0}, \quad (29)$$

where P_0 is the atmospheric pressure, γ is the specific heat ratio of the air, and H' is the thermal characteristic frequency of the poroelastic medium which is related to the thermal characteristic dimension Λ' introduced by Champoux and Allard.⁴ At low frequencies, it is noted that the real part of the bulk modulus may be considered constant, while its imaginary part varies linearly as a function of the frequency. Once again, Eq. (27) is obtained by keeping the first-order term of the series expansion of $K_f(\omega)$ in terms of the thermal dimensionless frequency ω/H' . It is solely valid for $\omega/H' \ll 1$. For frequencies approaching the thermal transition frequency ω_t , where $\text{Im}[K_f(\omega)]$ reaches its maximum, the linear increase of the imaginary part does not hold anymore. In this case, a medium-frequency approximation can be used. It consists of using a constant imaginary part, $\bar{\Gamma}$, in Eq. (27). In the upcoming results, $\bar{\Gamma} = \Gamma(0.4 \cdot \omega_t)$. This value minimizes the approximation error on $\text{Im}[K_f(\omega)]$ in the medium frequency range.²⁴ The low- and medium-frequency approximations on the imaginary part of the bulk modulus and its “exact” value are given in Fig. 2. The “exact” mathematical expressions of $K_f(\omega)$ in the Biot–Allard model can be found in Ref. 18.

Using the aforementioned low-frequency approximations, Eq. (23) can be rewritten as

$$(-\omega^2[\hat{M}_p] + j\omega(\beta[\hat{C}_p] + (P_0\Gamma(\omega)/\omega)[K_2]) + [\hat{K}_p])\{\bar{W}_p\} = \{F_p\}, \quad (30)$$

where $[\hat{K}_p] = [K_1] + P_0[K_2]$, and $[\hat{M}_p]$ is the “mass” matrix given in Eq. (15) in which the Biot’s densities are modified to take into account of the viscous effects. The modified Biot’s densities are given by

$$\rho'_{ij} = \rho_{ij} + (-1)^{i+j}\alpha, \quad (31)$$

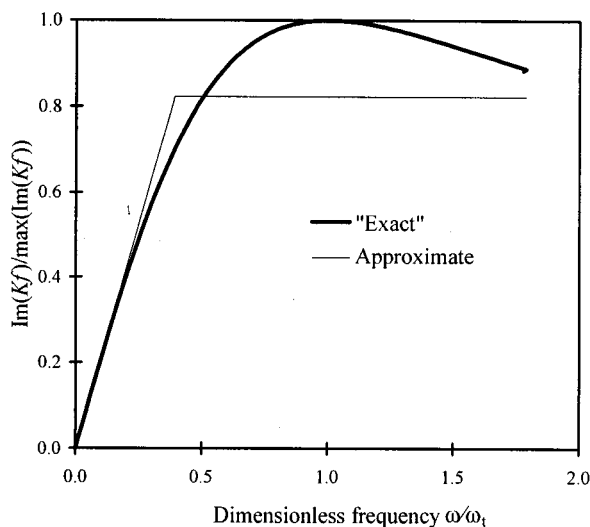


FIG. 2. Normalized imaginary part of the bulk modulus as a function of the frequency scaled to the thermal transition frequency.

where $i = 1, 2$ and $j = 1, 2$. It is noted that the imaginary part α of the viscous damping coefficient acts as an added density. In the case where the medium-frequency approximation is used for the bulk modulus, the term $(P_0 \Gamma(\omega)/\omega)[K_2]$ is removed from Eq. (30), and $[\hat{K}_p] = [K_1] + P_0(1 + j\Gamma)[K_2]$. To prove the accuracy of the efficient algorithm leading to Eq. (30) with the aforementioned approximations, Eq. (30) will be compared to “exact” analytical results in the next section. Other validations of the aforementioned approximations can be found in Ref. 24.

The advantage of using Eq. (30) is twofold. Firstly, it authorizes the use of classical FEM codes to handle poroelastic media. In fact, Eq. (30) uses real densities, a real viscous damping coefficient, and frequency-independent elasticity coefficients. Therefore, the matrices related to the solid and fluid phases can be calculated using solid elements provided in classical FEM codes. Secondly, Eq. (30) is more appropriate for conventional modal analysis methods. In fact, the low-frequency approximations linearize the related eigenvalue problem. Thus, it is believed that this efficient algorithm may be applied to solve the forced response of a poroelastic material using a modal analysis technique. This latter advantage will not be used in the forthcoming numerical simulations. Future works will investigate its application.

VII. RESULTS

In the following, the developed finite element Biot–Allard model is validated by comparisons with existing literature for both two-dimensional and three-dimensional problems. Also, the edge effect on the acoustic behavior of a multilayered poroelastic material are investigated. For the first three examples, the numerical results are obtained using Eq. (30) with the low-frequency approximations, and with Eq. (23) without the approximations. This is done in order to prove the validity of the low-frequency approximations and of the finite element model. For the last example, the numeri-

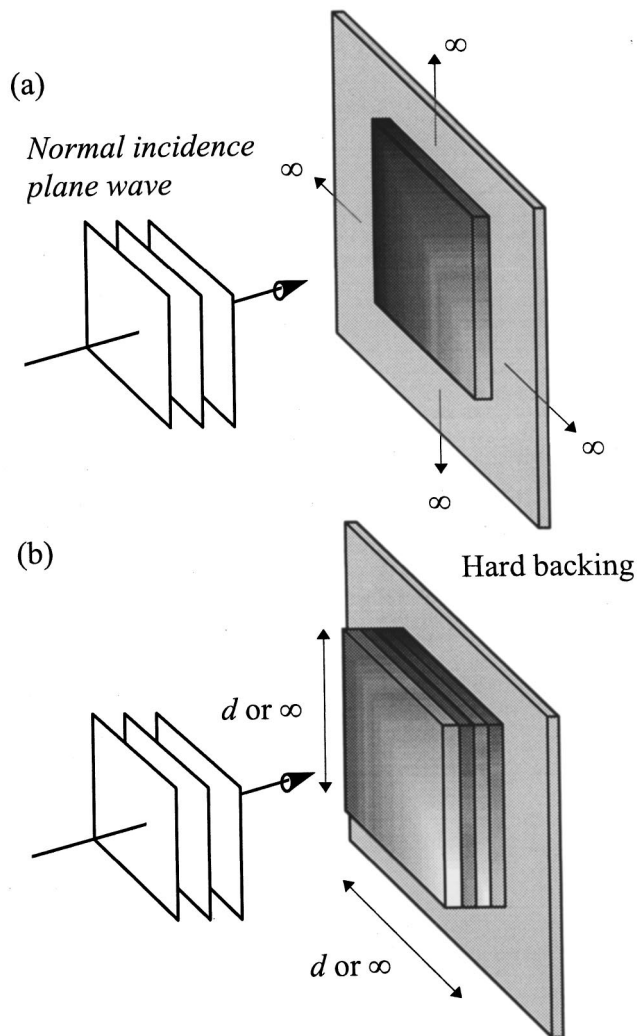


FIG. 3. Geometry of the surface impedance problem. (a) single poroelastic layer, and (b) multilayered poroelastic material.

cal results are obtained with Eq. (30) with the approximations. A direct solution method is used to solve the matrix systems.

A. Surface impedance for a single poroelastic layer

As a first example, the surface impedance of laterally infinite poroelastic materials predicted by the finite element Biot–Allard model with and without the low-frequency approximations is compared with an “exact” analytical calculation by Allard.¹⁸ The configuration, under study, is depicted in Fig. 3(a). A laterally infinite glass wool layer, described in Table I, is bonded onto a rigid impervious wall. A normal incidence plane wave of unit amplitude excites the absorbing material. To simulate the laterally infinite extent with the finite element model, only the axial macroscopic displacements are considered for the equivalent one-dimensional mesh depicted in Fig. 4., i.e., the lateral displacements are set to zero. The poroelastic domain is meshed using a six-node linear pentahedron element with six degrees of freedom per node: the three solid and the three fluid macroscopic displacements. The normal incidence surface impedance is calculated using the axial nodal solid and fluid

TABLE I. Physical properties and dimensions of the poroelastic materials.

| | k_s | ρ_1^a kg/m ³ | σ kNs/m ⁴ | h | N (N/cm ²) | ν_b^b | Λ (μ m) | Λ' (μ m) | H (Hz) | H' (Hz) | Thickness (cm) |
|--|-------|---------------------------------|--------------------------------|------|-----------------------------|----------------|-------------------------|--------------------------|-------------|--------------|-------------------|
| <i>Single poroelastic layer</i> | | | | | | | | | | | |
| Glass wool | 1.06 | 130 | 40 | 0.94 | 220(1 + j0.1) | 0 ^c | 56 | 110 | 7034 | 4496 | 10 |
| <i>Multilayered poroelastic material</i> | | | | | | | | | | | |
| Blanket | 1.18 | 41 | 34 | 0.98 | 11(1 + j0.015) | 0.3 | 60 | 87 | 5117 | 7188 | 0.4 |
| Screen | 2.56 | 125 | 3200 | 0.80 | 100(1 + j0.1) | 0.3 | 6 | 24 | 64177 | 94459 | 0.08 |
| Foam A | 2.52 | 31 | 87 | 0.97 | 5.5 (1 + j0.055) | 0.3 | 37 | 119 | 2736 | 3842 | 0.5 |
| Foam B | 1.98 | 16 | 65 | 0.99 | 1.8(1 + j0.1) | 0.3 | 37 | 121 | 2577 | 3715 | 1.6 |
| <i>Absorbing wall in the cavity</i> | | | | | | | | | | | |
| Foam C (60×60 cm ²) | 7.8 | 30 | 25 | 0.90 | 28.6 (1 + j0.27) | 0.4 | 226 | 226 | 757 | 1064 | 7.62 |

^aBulk density of the solid phase of the poroelastic material.

^bBulk Poisson's ratio of the poroelastic material.

^cZero Poisson's ratio is justified for fibrous materials with high porosity.²⁷

components, \bar{w} and \bar{W} , respectively, at the input surface for the unit acoustic pressure excitation by the following equation:

$$Z_n = \frac{1}{j\omega[h\bar{W} + (1-h)\bar{w}]} \quad (32)$$

Figure 5 presents the real and imaginary parts of the surface impedance predicted by the finite element Biot–Allard model, with and without the low-frequency approximations, and the analytical Biot–Allard model¹⁸ for a single poroelastic layer. To reach convergence in the (300–1200 Hz) frequency range, four linear elements were used in the axial direction. An excellent agreement between both models is obtained. Also, it is noted that the low-frequency approximations give good results. This is logical since the characteristic frequencies for the glass wool are much greater than 1200 Hz (see Table I). Hence, $\omega/H \ll 1$ and $\omega/H' \ll 1$ in the studied frequency range.

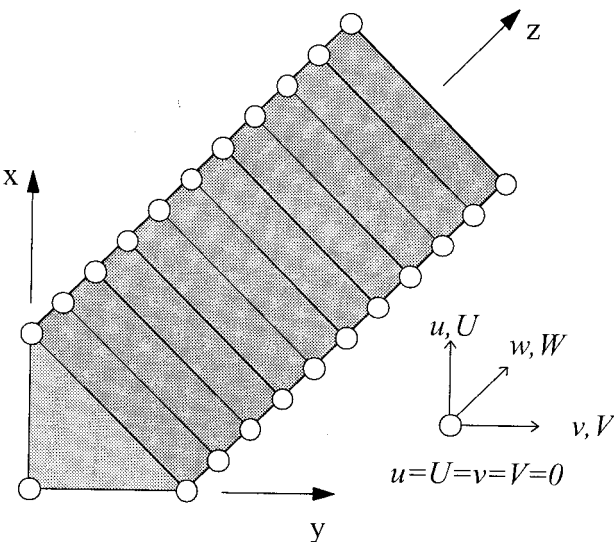


FIG. 4. One-dimensional finite element discretization.

B. Surface impedance for a multilayered poroelastic material

To check the ability of the model to handle multilayered poroelastic materials, we redo the calculation of the previous example for the configuration depicted in Fig. 3(b). From the front face to the rear face, the multilayer consists of a blanket, a screen, foam A, and foam B. The material properties are listed in Table I. A comparison is made with an analytical calculation by Allard¹⁸ for the (100–4000 Hz) frequency range. The results are shown in Fig. 6. For the results shown, the blanket, the screen, and foam A were discretized using only one element, and foam B with two elements. Again, an excellent agreement between both models is obtained and the low-frequency approximations give acceptable results for the studied frequency range. However, for frequencies greater than 2000 Hz, the error introduced by the low-frequency

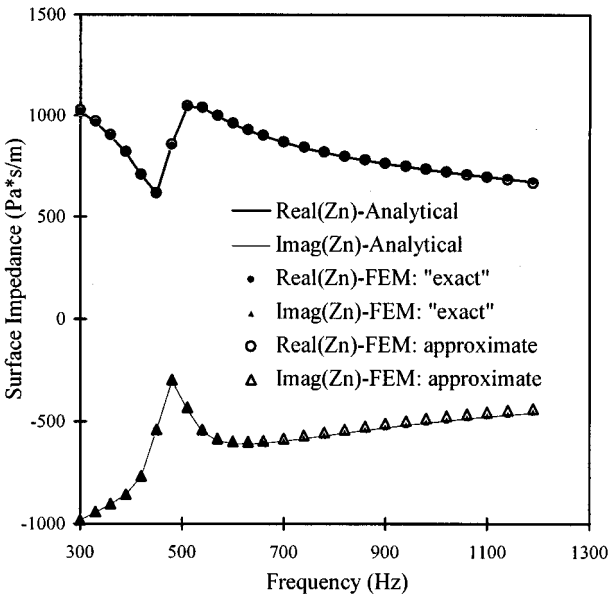


FIG. 5. Comparison of the surface impedance between an analytical model¹⁸ and the finite element model for the laterally infinite glass wool layer of Table I.

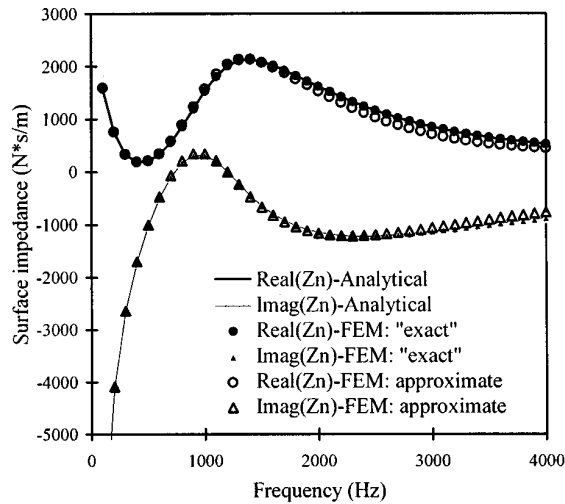


FIG. 6. Comparison of the surface impedance between an analytical model¹⁸ and the finite element model for the laterally infinite multilayered poroelastic material of Table I.

approximations is more important but still acceptable. This is due to the fact that the characteristic frequencies for the foam layers are in the vicinity of the upper limit of the studied frequency range (see Table I).

The excellent results of the two previous validation examples prove the validity of the proposed efficient algorithm leading to Eq. (30) with respect to “exact” analytical calculations. Moreover, since the analytical calculations were validated with experimental measurements,¹⁸ the presented finite element Biot–Allard model is indirectly experimentally validated.

C. Interior three-dimensional problem

To validate the developed model for a three-dimensional case, let us consider the configuration depicted in Fig. 7. It consists of a rigid cubic enclosure of 1 m³. A volume displacement of 2.5×10^{-8} m³ is imposed at $(x, y, z) = (0, 0, 0)$. A partial acoustic treatment is applied to one wall. It consists of a 60×60 -cm² porous material embedded into the wall. All its faces but the front face are constrained. The material used is foam C given in Table I. The structural damping coefficient for the air in the enclosure is $\eta_a = 0.001$.

The cavity is discretized using six-node linear pentahedron elements with one degree of freedom per node: the acoustic pressure. A $10 \times 10 \times 10$ mesh is used. In the absence of the foam, this mesh has been found satisfactory for the frequency range of interest by comparing the numerically computed eigenfrequencies of the cavity to their analytical values. For the poroelastic domain, an $8 \times 8 \times 6$ mesh is used. The mean quadratic pressure in the enclosure is used as the indicator for this test. It is defined by

$$\langle P^2 \rangle = \frac{1}{2N_a} \sum_{i=1}^{N_a} |\bar{P}_i|^2, \quad (33)$$

where N_a is the number of nodes in the discretized acoustic domain. In decibel (dB), the associated level is

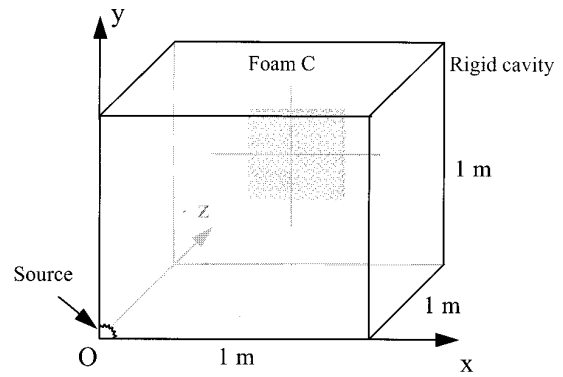


FIG. 7. Geometry of the 3-D rigid cavity problem.

$$L_P = 10 \log \left(\frac{\langle P^2 \rangle}{4 \times 10^{-10}} \right). \quad (34)$$

The results obtained with the proposed coupled finite element model of Eq. (22) with and without the low-frequency approximations are compared with an admittance approach for the poroelastic material. The admittance approach uses the acoustic finite element model of Eq. (16) in which an admittance matrix is added. The admittance matrix is given by²²

$$[A] = j \frac{\omega}{c_0} \sum_{e \in \Sigma_{pa}} \int_{\Sigma_{pa}^e} \{N_p\}^e \beta \{N_p\}^{e^T} dS, \quad (35)$$

where β is the admittance of the absorbing surface Σ_{pa} . It is given by $\rho_0 c_0 / Z_n$ where Z_n is given by the “exact” analytical Biot–Allard model of Allard.¹⁸ The results of the comparison are shown in Fig. 8. An excellent agreement between the two methods is obtained. Once again, the low-frequency approximations work well. This test confirms both the validity of the proposed efficient algorithm and its usefulness to handle a practical configuration. It is worth noting that, in comparison with the proposed finite element model, the admittance approach takes a much smaller solution time. In-

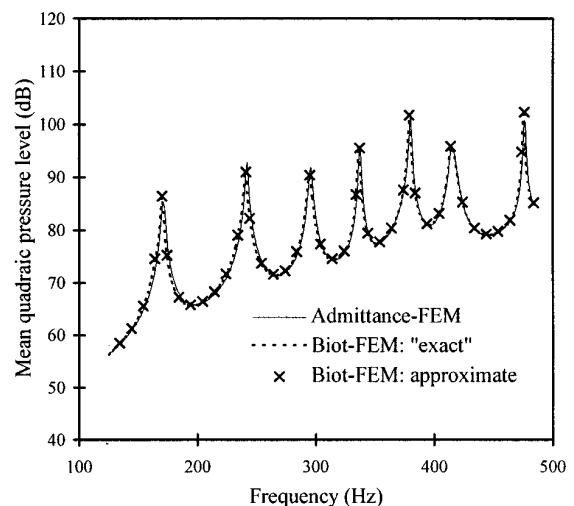


FIG. 8. Mean quadratic pressure inside the rigid cavity with partial treatment with foam C of Table I. Comparison between the Biot–Allard approach and the admittance approach.

deed, with the admittance approach, only the cavity needs to be discretized with acoustic elements. However, this approach is not suited to the case where the wall, backing the poroelastic material, is flexible; such as the roof of an automobile cabin. In contrast, the proposed model can handle such situations upon its coupling with elastic domains.^{25,26}

D. Edge effect on the absorption

To show the usefulness of the model, let us consider the edge effect on the absorption of the multilayered poroelastic material shown in Fig. 3(b) and described in Table I. This time, the multilayer is of finite extent. A square sample of width d is chosen. The edge effect will be regarded for different widths. The sample has two constrained edges and two lubricated edges. The constrained edges imply that all but the axial fluid displacements vanish on the respective edges while the lubricated edges only allow for axial displacements. Figure 9 compares the absorption coefficients of the constrained multilayer for 3 different widths: $d=2.5$, 5, and 10 cm. Also, these results are compared to the absorption of the laterally infinite multilayer. Here, the numerical results obtained with Eq. (29) and the approximations are used to evaluate the absorption coefficient. This coefficient is given by¹⁸

$$\alpha = 1 - \left| \frac{Z_n - \rho_0 c_0}{Z_n + \rho_0 c_0} \right|^2, \quad (36)$$

where Z_n is given by Eq. (32). In Fig. 9, it is noted that the edge effect modifies the absorption coefficient of the multilayered poroelastic material in comparison with the laterally infinite sample. However, as the width increases, the behavior approaches that of the laterally infinite sample. Note also that as the frequency increases the absorption coefficient approaches that of the laterally infinite sample. This is a classical result explaining why commonly used analytical models based on the laterally infinite layer assumption give good results at high frequencies. However, at low frequencies, the acoustical behavior of a poroelastic sample is sensitive to the boundary conditions along its edges. Consequently, these analytical models and also acoustical measurements in a standing wave tube may be misleading if not properly interpreted. Similar results can be found in Ref. 13.

Finally, it is worth mentioning that for the sample with $d = 2.5$ cm, 15 poroelastic linear wedge elements along the width of the material were necessary to reach convergence, i.e., to adequately evaluate the shear waves. This is likely due to the fact that the linear wedge element is not efficient in bending. In comparison, only 5 elements along the thickness (total thickness is 2.58 cm) were necessary to adequately evaluate the compression waves. We have found that a simple meshing criterion, such as typically used for elastic and fluid domains, may be misleading for poroelastic domains due to the complex coupling between the solid phase and the fluid phase. Further analyses, particularly based on a modal analysis method, are needed to provide a meshing criterion for poroelastic materials.

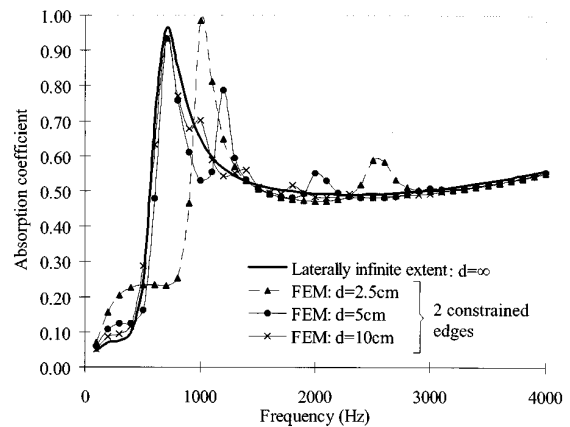


FIG. 9. Edge effect on the absorption coefficient for the multilayered poroelastic material of Table I. Two edges are constrained and two are lubricated.

VIII. CONCLUSION

In this paper, a three-dimensional finite element approach has been used to solve the poroelasticity problem in acoustics based on the isotropic Biot–Allard theory. A displacement finite element model has been derived using the Lagrangian approach together with an analogy with solid elements. It has been shown that the corresponding discretized equations of motion lead to a symmetrical matrix system in which the solid and fluid degrees of freedom are explicitly separated. Moreover, it has been noted that the “damping” and “stiffness” matrices of the poroelastic medium are complex and frequency dependent. This leads to cumbersome calculations for large finite element models and spectral analyses since these matrices are huge and need to be calculated at each frequency step. To overcome this difficulty and to boost the efficiency of the algorithm, the discretized equations of motion have been transformed to a more appropriate form; using the fact that the solid and fluid degrees of freedom are explicitly separated. In this appropriate form, the frequency-dependent coefficients were taken out from the “damping” and “stiffness” matrices. This leads to a reduction in the number of calculations at each frequency step. However, for large finite element models, the solution time is still prohibitive.

To further simplify the equations of motion, realistic low-frequency approximations have been made on the frequency-dependent viscous damping coefficient and the bulk modulus of the air in the pores. It has been shown that these approximations enable the use of classical FEM codes to model poroelastic materials. Also, the low-frequency approximations linearize the poroelastic eigenvalue problem. It is believed that the resulting efficient algorithm may be applied to solve the forced response of a poroelastic material using a modal analysis technique. Works using modal analysis techniques on the discretized poroelastic equations with the low-frequency approximations are underway.

The efficient finite element Biot–Allard model has been compared to existing literature for both two-dimensional and three-dimensional problems. Excellent comparisons have

proved both the accuracy and effectiveness of the finite element Biot–Allard model, with and without the low-frequency approximations, and its coupling with acoustic elements. Also, to show the usefulness of the model, the edge effect on the absorption coefficient of a multilayered poroelastic material has been presented. It has been shown that constraining the edges may affect considerably the absorption coefficient at low frequencies, especially for small samples. Nevertheless, for larger samples or at higher frequencies, the edge effect is no more noticeable. This is a classical result explaining why commonly used analytical models, based on laterally infinite poroelastic layers, accurately predict the high-frequency behavior. However, at low frequencies, these analytical models, and also standing wave tube measurements may be misleadingly interpreted.

Finally, the presented three-dimensional finite element model can be used in various coupled problems of complex geometries.^{25,26} Using appropriate coupling conditions, it is compatible with classical elastic finite elements. However, it should be pointed out that the authors believe that the development of efficient solution methods, in the sense of the one described in this paper, is the key for handling realistic spectral analyses on medium to large finite element models. Consequently this should be a concern for future works.

ACKNOWLEDGMENTS

The authors wish to thank Professor Jean Nicolas for his help and support during this work and Professor J.-F. Allard for his contribution to the validation of the proposed model. Thanks are also expressed to Bombardier Inc. Canada, N.S.E.R.C. (Natural Sciences and Engineering Research Council of Canada), and F.C.A.R. (Fonds pour la Formation de Chercheurs et l'Aide à la Recherche) for their financial support.

APPENDIX

1. Spatial derivative operator

$$[L]^T = \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 & \frac{\partial}{\partial y} & 0 & \frac{\partial}{\partial z} \\ 0 & \frac{\partial}{\partial y} & 0 & \frac{\partial}{\partial x} & \frac{\partial}{\partial z} & 0 \\ 0 & 0 & \frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{bmatrix}.$$

2. Poroelastic coefficient A for highly porous materials

$$A \cong 2N \frac{\nu_b}{(1 - 2\nu_b)}.$$

3. Direction cosines matrix

$$[N]^T = \begin{bmatrix} n_x & 0 & 0 & n_y & 0 & n_z \\ 0 & n_y & 0 & n_x & n_z & 0 \\ 0 & 0 & n_z & 0 & n_y & n_x \end{bmatrix},$$

where n_x , n_y , and n_z are the components of the unit outward normal vector $\{n\}$ to the acoustic medium.

- ¹ C. Zwikker and C. W. Kosten, *Sound-Absorbing Materials* (Elsevier, New York, 1949).
- ² P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1978).
- ³ K. Attenborough, "Acoustical characteristics of rigid fibrous absorbents and granular media," *J. Acoust. Soc. Am.* **73**, 785–799 (1983).
- ⁴ Y. Champoux and J.-F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," *J. Appl. Phys.* **70**, 1975–1979 (1991).
- ⁵ M. A. Biot, *Acoustics, Elasticity, and Thermodynamics of Porous Media. Twenty-one Papers by M. A. Biot*, edited by I. Tolstoy (Acoustical Society of America, Woodbury, NY, 1992).
- ⁶ M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low frequency range," *J. Acoust. Soc. Am.* **28**, 168–191 (1956).
- ⁷ J.-F. Allard, A. Aknine, and C. Depollier, "Acoustical properties of partially reticulated foams with high and medium flow resistance," *J. Acoust. Soc. Am.* **79**, 1734–1740 (1986).
- ⁸ N.-M. Shiau, "Multi-dimensional wave propagation in elastic porous materials," Ph.D. thesis, School of Mechanical Engineering, Purdue University, 1991.
- ⁹ J. S. Bolton and E. R. Green, "Normal incidence sound transmission through double panel systems lined with relatively stiff, partially reticulated polyurethane foam," *Appl. Acoust.* **39**, 23–51 (1993).
- ¹⁰ A. Craggs, "A finite element model for rigid porous absorbing materials," *J. Sound Vib.* **61**, 101–111 (1978).
- ¹¹ R. Panneton, N. Atalla, and F. Charon, "A finite element formulation for the vibro-acoustic behaviour of double plate structures with cavity absorption," *Can. Aero. Space J.* **41**, 5–12 (1995).
- ¹² R. Panneton, N. Atalla, and J.-F. Allard, "A three-dimensional finite element model for sound transmission through a double-plate system with isotropic elastic porous material," *J. Acoust. Soc. Am.* **96**, 3339A (1994).
- ¹³ Y. J. Kang and J. S. Bolton, "Finite element modeling of isotropic elastic porous materials coupled with acoustical finite elements," *J. Acoust. Soc. Am.* **98**, 635–643 (1995).
- ¹⁴ T. F. Johansen, J.-F. Allard, and B. Brouard, "Finite element method for predicting the acoustical properties of porous samples," *Acta Acust.* (submitted).
- ¹⁵ J. P. Coyette and H. Wynendaele, "A finite element model for predicting the acoustic transmission characteristics of layered structures," in *Proceedings of Inter-Noise 95*, edited by R. J. Bernhard and J. S. Bolton (Noise Control Foundation, Poughkeepsie, NY, 1995), pp. 1279–1282.
- ¹⁶ O. C. Zienkiewicz, *The Finite Element Method in Engineering Science* (McGraw-Hill, New York, 1972).
- ¹⁷ P. Göransson, "A weighted residual formulation of the acoustic wave propagation through a flexible porous material and comparison with a limp material model," *J. Sound Vib.* **182**, 479–494 (1995).
- ¹⁸ J.-F. Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Elsevier, New York, 1993).
- ¹⁹ D. L. Johnson, J. Koplik, and L. M. Schwartz, "New pore-size parameter characterizing transport in porous media," *Phys. Rev. Lett.* **57**, 2564–1979 (1986).
- ²⁰ B. R. Simon, J. S. S. Wu, O. C. Zienkiewicz, and D. K. Paul, "Evaluation of $u-w$ and $u-\pi$ finite element formulation for the dynamic response of saturated porous media using one-dimensional models," *Int. J. Numer. Anal. Meth. Geomech.* **10**, 461–482 (1986).
- ²¹ W. C. Hurty and M. F. Rubinstein, *Dynamics of Structures* (Prentice-Hall, Englewood Cliffs, NJ, 1964).
- ²² C. Lesueur, *Rayonnement acoustique des structures* (Eyrolles, Paris, 1988), pp. 355–406.
- ²³ O. M. Lovera, "Boundary conditions for a fluid-saturated porous solid," *Geophysics* **52**, 174–178 (1987).
- ²⁴ R. Panneton and N. Atalla, "Modélisation numérique 3D par éléments finis des milieux poroélastiques: application au problème couplé élasto-poro-acoustique," Ph.D. thesis, Mech. Eng. Dept., Université de Sherbrooke, 1996.

- ²⁵R. Panneton and N. Atalla, "Numerical prediction of sound transmission through multilayer systems with poroelastic materials," *J. Acoust. Soc. Am.* **100**, 346–354 (1996).
- ²⁶N. Atalla and R. Panneton, "The effects of multilayer sound-absorbing treatments on the noise field inside a plate backed cavity," *Noise Control Eng. J.* **44**, 235–243 (1996).
- ²⁷D. J. Sides, K. Attenborough, and K. A. Mulholland, "Application of a generalized acoustic propagation theory of fibrous absorbents," *J. Sound Vib.* **19**, 49–64 (1971).
- ²⁸H. Deseriewicz and R. Skalak, "On uniqueness in dynamic poroelasticity," *Bull. Seism. Soc. Am.* **53**, 783–788 (1963).

Acoustic dispersion and attenuation in many spherical scatterer systems and the Kramers–Kronig relations

Zhen Ye^{a)}

Institute of Ocean Sciences, Sidney, British Columbia, Canada and

Department of Physics, National Central University, Chung Li, Taiwan, Republic of China

(Received 17 September 1996; revised 15 December 1996; accepted 6 January 1997)

This paper presents a theoretical study of the acoustic dispersion in media containing many spherical scatterers. The existing dispersion relation is used to investigate the dispersion in media containing particles such as low acoustic-contrast fluid, rigid and elastic spheres, as well as air bubbles. It is shown that in the case of low acoustic-contrast fluid, rigid and elastic spheres, the presence of scatterers has more profound effects on the acoustic attenuation than on the phase speed, whereas in the bubble case, both the attenuation and phase speed can be affected significantly. Then the dispersion relation is examined via the Kramers–Kronig relations, as they express a general identity for dispersion and are therefore ideal to test the suitability of a specific model for acoustic propagation. Results indicate that although there remain some discrepancies, the existing dispersion relation satisfies the Kramers–Kronig relations reasonably well. Good agreement is particularly apparent in the bubble case. Possible cause for the discrepancies is also discussed in connection with a discussion of the assumptions that lead to the dispersion relation. © 1997 Acoustical Society of America. [S0001-4966(97)02505-8]

PACS numbers: 43.20.Gp, 43.20.Fn, 43.30.Xm [JEG]

INTRODUCTION

When waves propagate through a medium containing many randomly distributed scatterers, they are multiply scattered, resulting in that the phase speed varies with frequency. This phenomenon is called dispersion. The problem of multiple scattering of acoustic waves by a random distribution of scatterers has been studied intensively in the past, not only because of their vast applications to such as scattering by underwater air bubbles,¹ fog,² fish schools,³ sand,⁴ and turbulence,⁵ but also because the problem itself is theoretically challenging.

The work of Foldy,⁶ Lax,⁷ Waterman and Truell,⁸ and Twersky⁹ serves as a foundation for the study of multiple scattering. Foldy⁶ first considered the multiple scattering of scalar waves by a random distribution of isotropic scatterers on the basis of a self-consistent method. He introduced the concept of “randomness” which requires average to be taken over a statistical ensemble of scatterer configurations. Under certain approximations to be discussed later, Foldy derived a classic dispersion relation for the coherent wave propagation through a medium containing many isotropic scatterers of the same size,

$$\kappa^2 = k^2 + 4\pi n f, \quad (1)$$

where κ is the effective wave number in the presence of the many scatterers, k is the wave number when the scatterers are absent, f is the isotropic scattering function, and n is the numerical density of the scatterers (i.e., number/m³). The scattering function will be discussed in more detail later. This Foldy result has been used widely for studies of low-frequency sound propagation in water containing many inter-

acting air bubbles (e.g., Refs. 1, 10), for which the scattering function can be approximated as isotropic.

Later, Lax⁷ extended the work of Foldy to include anisotropic scatterers. He considered wave propagation through a medium containing many spherical scatterers of the same size. The dispersion relation derived by Lax is written as

$$\kappa^2 = k^2 + 4\pi n f(\hat{k}_i, \hat{k}_i), \quad (2)$$

where $f(\hat{k}_i, \hat{k}_i)$ is the scattering function in the forward direction. The result of Lax was further improved by Waterman and Truell,⁸ who took into account the contribution from backscattering more rigorously. The dispersion relation of Waterman and Truell is written

$$\kappa^2 = k^2 \left(\left[1 + \frac{2\pi n f(\hat{k}_i, \hat{k}_i)}{k^2} \right]^2 - \left[\frac{2\pi n f(\hat{k}_i, -\hat{k}_i)}{k^2} \right]^2 \right), \quad (3)$$

where $f(\hat{k}_i, -\hat{k}_i)$ is the backscattering function. When the scatterer concentration is low, so that the second-order effects associated with n^2 can be ignored, the dispersion relation of Waterman and Truell reduces to that of Lax. Furthermore, if the scattering function is isotropic, the results of Lax and Waterman reduce to that of Foldy. Therefore, the results of Lax and Foldy are two special cases of Waterman and Truell's, and thus we need only consider the relation derived by Waterman and Truell.

The purpose of this paper is to present a study of the acoustic dispersion in a medium containing many scatterers, using the above dispersion formulas. Since they are approximate, the validity of these dispersion relations deserves examination. For this purpose, the Kramers–Kronig (KK) relations^{11,12} will be used. This is because the Kramers–Kronig relations express a general relation for wave dispersion that is independent of specific models. They are, there-

^{a)}Electronic mail: zhen@joule.phy.ncu.edu.tw

fore, ideal to test the suitability of a specific model for acoustic propagation. Particularly, we will consider acoustic dispersion in water containing fluid spheres, elastic spheres, and air-filled bubbles. Fluid spheres may be used to model oil spills, zooplanktons, and fogs, whereas elastic spheres and air bubbles may model sandy suspensions and fish swim-bladders, respectively.

I. THE ANALYSIS

A. Dispersion relation and the Kramers–Kronig relations

The concept of dispersion relations was introduced to physics by the work of Kramers and Kronig;¹¹ the dispersion is a result of the dependence of the index of refraction on wavelength or angular frequency. The real part of the index of refraction (n_r) is determined by the phase velocity, while the imaginary part is determined by the absorption. Kramers and Kronig showed that the real part of $n_r^2 - 1$ can be expressed as an integral of the imaginary part, and vice versa:¹³

$$\text{Re}[n_r^2(\omega) - 1] = \frac{2}{\pi} \mathcal{P} \int_0^\infty ds \frac{s \text{Im}[n_r^2(s)]}{s^2 - \omega^2}, \quad (4)$$

and

$$\text{Im}[n_r^2(\omega)] = -\frac{2\omega}{\pi} \mathcal{P} \int_0^\infty ds \frac{\text{Re}[n_r^2(s) - 1]}{s^2 - \omega^2}, \quad (5)$$

where \mathcal{P} denotes the principal value. These relations are called Kramers–Kronig (KK) relations, and the integrals in these relations are referred to as the Hilbert transformation. Since they express a general relation as a result of causality and are independent of the details of particular interactions, these relations can be therefore used to examine any particular model for the refractive index.

In the problem of acoustic wave propagation in a medium containing many randomly distributed scatterers, the wave equation can be generally written as¹⁴

$$(\nabla^2 + k^2)p = J_s[p] - 4\pi\delta^{(3)}(\mathbf{r} - \mathbf{r}_s), \quad (6)$$

where $J_s[p]$ refers to interaction terms, generically representing the interaction between the scatterers and the propagating wave, and $\delta^{(3)}(\cdot)$ refers to the source term. An iterative solution to Eq. (6) can be obtained using, for example, the diagram rules.¹⁴ The resulting averaged wave equation, known as the Dyson equation, can be written as

$$(\nabla^2 + \kappa^2)\langle p \rangle = -4\pi\delta^{(3)}(\mathbf{r} - \mathbf{r}_s), \quad (7)$$

where the angle brackets refer to an ensemble average over the scatterer distribution.⁶ Here κ is identified as the effective wave number, and is a function of the wave number when no scatterers are present. In this article, we will use the KK relations to inspect the dispersion relation derived by Waterman and Truell.⁸ We repeatedly write it out as follows,

$$\kappa^2 = k^2 \left(\left[1 + \frac{2\pi n f(\hat{k}_i, \hat{k}_i)}{k^2} \right]^2 - \left[\frac{2\pi n f(\hat{k}_i, -\hat{k}_i)}{k^2} \right]^2 \right).$$

This yields the index of refraction as,

$$n_r^2 = \left[1 + \frac{2\pi n f(\hat{k}_i, \hat{k}_i)}{k^2} \right]^2 - \left[\frac{2\pi n f(\hat{k}_i, -\hat{k}_i)}{k^2} \right]^2. \quad (8)$$

For low scatterer concentrations and when the backscattering is negligible compared to the forward scattering, this can be approximated as

$$n_r^2 \approx 1 + \frac{4\pi n f(\hat{k}_i, \hat{k}_i)}{k^2}, \quad (9)$$

or, for isotropic scatterers, Eq. (8) becomes

$$n_r^2 = 1 + \frac{4\pi n f}{k^2}. \quad (10)$$

We will examine to what extent the refractive index in Eq. (8) satisfies the KK relations in Eqs. (4) and (5).

Once the effective wave number is determined, the phase speed (v) and the acoustic attenuation in dB per unit distance (α) can be determined from

$$v = \frac{\omega}{\text{Re}(\kappa)}, \quad \alpha = 8.68 \text{Im}(\kappa), \quad (11)$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ mean taking the real and the imaginary part, respectively. Both v and α are measurable quantities. For low scatterer concentrations and when the sound-speed anomaly is much smaller than the sound speed in water, i.e., $\Delta c \ll c_w$, where c_w is the sound speed in water, the first expression of Eq. (11) can be approximately written as

$$\Delta c \approx -\text{Re} \left(\frac{2\pi n f(\hat{k}_i, \hat{k}_i)}{k^2} c_w \right), \quad (12)$$

where $\Delta c = v - c_w$, with v being the effective phase speed in the medium. The above equation indicates that the phase-speed anomaly in a medium containing many scatterers of equal size is approximately proportional to the scattering function of a single scatterer. It is interesting to note from Eq. (9) that at low scatterer concentrations and when the backscattering is weak, the KK relations for the dispersion relation reduces to the following identities for the forward scattering function of a single scatterer, i.e.,

$$\text{Re}[f(\hat{k}_i, \hat{k}_i; \omega)/\omega^2] = \frac{2}{\pi} \mathcal{P} \int_0^\infty ds \frac{\text{Im}[f(\hat{k}_i, \hat{k}_i; s)]}{s(s^2 - \omega^2)}, \quad (13)$$

and

$$\text{Im}[f(\hat{k}_i, \hat{k}_i; \omega)/\omega^2] = -\frac{2\omega}{\pi} \mathcal{P} \int_0^\infty ds \frac{\text{Re}[f(\hat{k}_i, \hat{k}_i; s)]}{s^2(s^2 - \omega^2)}. \quad (14)$$

In other words, the validity of the dispersion relation for wave propagation in many scatterers can be examined through the properties of scattering by a single scatterer.

B. Scattering functions

It is instructive to note that the behavior of the scattering medium, characterized by the complex wave number κ , is specified explicitly in terms of the number of scatterers per

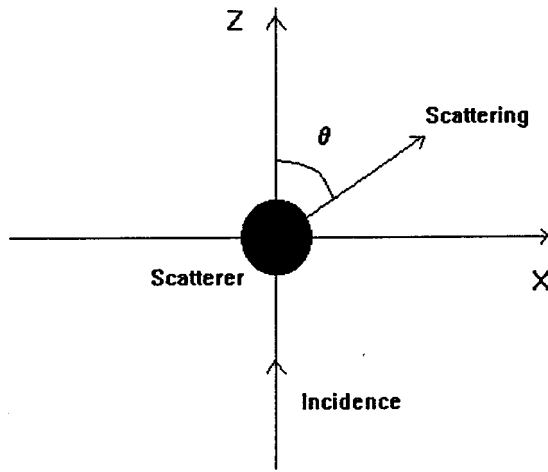


FIG. 1. Geometry for sound scattering a spherical scatterer.

unit volume and the far-field scattering function for a single scatterer. We first look at the scattering function.

Consider a unit plane wave, $\exp(ik\hat{k}_i \cdot \mathbf{r})$, incident on a scatterer located at the origin (see Fig. 1). Here \hat{k}_i is the unit wave vector denoting the incident direction. The scattered wave in the direction \hat{k}_s at far field can be written, according to the Sommerfeld radiation condition, as

$$p_s = f(\hat{k}_i, \hat{k}_s) \frac{e^{ikr}}{r}. \quad (15)$$

The scattering function $f(\hat{k}_i, \hat{k}_s)$ can be solved in terms of a modal series for the spherical scatterers under consideration.

For fluid spheres including air bubbles, the scattering function can be calculated as¹⁵

$$f(\hat{k}_i, \hat{k}_s) = \frac{i}{k} \sum_{l=0}^{\infty} (2l+1) B_l P_l(\cos \theta), \quad (16)$$

where

$$B_l = \frac{(-1)^l}{1 + iC_l},$$

with

$$C_l = \frac{n_l(ka)j'_l(k_1a) - ghj_l(k_1a)n'_l(ka)}{j_l(ka)j'_l(k_1a) - ghj_l(k_1a)j'_l(ka)},$$

in which k_1 denotes the wave number inside the scatterer. In the above P_l is the Legendre polynomial, j_l the spherical Bessel function of the first kind, n_l the spherical Neuman

TABLE I. Parameters used in the computation.

| Material | Mass density (kg/m ³) | Sound speed (m/s) | Shear wave speed (m/s) |
|-----------------------------|--------------------------------------|----------------------|---------------------------|
| Water (medium) | 1000 | 1500 | |
| Fluid sphere | 1040 | 1560 | |
| Rigid fluid sphere | ∞ | ∞ | |
| Elastic sphere ^a | 7900 | 6860 | 4185 |
| Air bubbles | 1.29 | 340 | |

^aStainless-steel sphere.

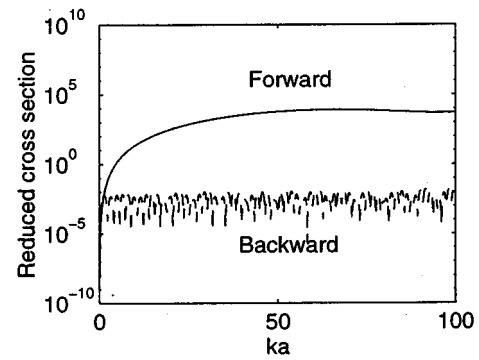


FIG. 2. Reduced scattering cross section versus ka for a low acoustic-constrast fluid sphere.

function, g and h are the density and speed ratios between the scatterer and the medium, respectively, and θ is the scattering angle given by $\theta = \cos^{-1}(\hat{k}_i \cdot \hat{k}_s)$. Throughout the paper, we use the following conventions

$$j'_l(x) = \frac{dj_l(x)}{dx},$$

and

$$n'_l(x) = \frac{dn_l(x)}{dx}.$$

The forward scattering and backscattering functions can be obtained by letting $\theta=0$ and π , respectively.

For an elastic sphere, the scattering function was originally derived by Faran,¹⁶ and later corrected by Hickling.¹⁷ It can be written as

$$f(\hat{k}_i, \hat{k}_s) = \frac{1}{k} \sum_{l=0}^{\infty} Q_l (-i)^{l+1} P_l(\cos \theta). \quad (17)$$

The coefficients Q_l are determined from the appropriate boundary conditions at the surface of the sphere, and can be shown to be

$$Q_l = k(-1)^l(2l+1)h_l^{(1)}(ka) \sin \eta_l \exp(-i\eta_l),$$

where the angle η_l is given by

$$\tan \eta_l = -\frac{j_l(ka)F_l - j'_l(ka)}{n_l(ka)F_l - n'_l(ka)}$$

with

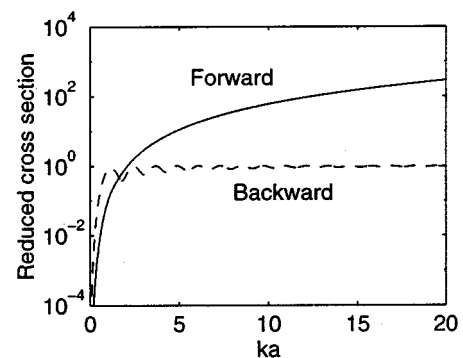


FIG. 3. Reduced scattering cross section versus ka for a rigid fluid sphere.

$$F_l = \frac{\rho}{\rho_1} \frac{x_2^2}{2}$$

$$\times \frac{[x_1 j_l'(x_1)]/[x_1 j_l'(x_1) - j_l(x_1)] - [2(l^2 + l)j_l(x_2)]/[(l^2 + l - 2)j_l(x_2) + x_2^2 j_l''(x_2)]}{[2x_1 j_l'(x_1) - j_l(x_1)(l^2 + l - x_2^2/2)]/[x_1 j_l'(x_1) - j_l(x_1)] - \{2(l^2 + l)[j_l(x_2) - x_2 j_l'(x_2)]/[(l^2 + l - 2)j_l(x_2) + x_2^2 j_l''(x_2)]\}}$$

and

$$x_1 = k_1 a, \quad x_2 = k_2 a.$$

In the above, ρ_1 is the mass density of the scatterer, k_1 , k_2 are the wave numbers for the compressional and shear waves of the solid sphere, respectively, and $h_l^{(1)}$ is the spherical Hankel function of the first kind, and

$$j_l''(x) \equiv \frac{d^2 j_l(x)}{dx^2}.$$

II. NUMERICAL RESULTS AND DISCUSSION

In this section we consider wave propagation through water containing many spherical scatterers with different material composition, i.e., rigid fluid, low acoustic-contrast fluid, elastic (stainless steel), and air. The parameters used in our computation are listed in Table I. In addition, the radius of the spheres is set at 1.0 mm.

We first examine the scattering functions. In Figs. 2–5, the reduced forward scattering and backscattering cross sections are plotted as a function of ka for the low acoustic-contrast fluid sphere, rigid sphere, elastic sphere (stainless steel), and air bubble, respectively. The reduced differential scattering cross section is defined as

$$\sigma(\theta) \equiv \frac{4}{a^2} |f(\theta)|^2. \quad (18)$$

From these figures, the following features follow. (1) In contrast to the backscattering, the forward scattering, except for the bubble case, varies rather smoothly with frequency and lacks the oscillatory features. (2) The forward scattering is significantly larger than the backscattering as the frequency increases, indicating the backscattering may be negligible at high frequencies. (3) There are small fine structures in the forward scattering for the elastic sphere case. (4) In the forward

direction, the scattering cross sections are of the same order of magnitude for low acoustic-contrast fluid and rigid and elastic spheres, whereas in the back direction, the rigid sphere does not have so many oscillatory features as the elastic sphere, but it has the same order of magnitude as the elastic sphere. (5) Figure 5 shows the usual resonant scattering feature in the bubble case;¹⁸ the scattering from the air bubble is isotropic in the frequency concerned ($ka \ll 0.1$), and the resonance appears at around $ka = 0.0136$ (Fig. 5). Note that in the bubble case, we have ignored the thermal and viscosity effects.¹⁸ When these effects are included, the resonance peak will be reduced considerably.

Now we consider the dispersion relations. To compute effective wave-number κ given by Eq. (3), a more practical parameter, volume fraction occupied by the scatterers, is used to replace n (number of particles m^3). For the spherical scatterers, the volume fraction is given by

$$\beta = \frac{4\pi}{3} a^3 n. \quad (19)$$

Substituting β into Eq. (12) leads to

$$\Delta c \approx -\beta \operatorname{Re} \left(\frac{3f_{\text{inf}}(\hat{k}_i, \hat{k}_i)}{(ka)^2} c_w \right), \quad (20)$$

where $f_{\text{inf}} = af(\hat{k}_i, \hat{k}_i)/2$ is the scattering form function and is dimensionless. Equation (20) shows that the phase-speed anomaly is proportional to the volume fraction.

Figures 6–9 present the plots of phase speed and attenuation as a function of ka using Eq. (3) in media containing low acoustic-contrast fluid spheres, rigid spheres, elastic spheres, and air filled spheres, respectively. The volume fraction of the scatterers is chosen as 0.001.

From these figures, we note the following. (1) The scatterers have much more pronounced effects on the attenuation than on the phase speed, for example, even with the rela-

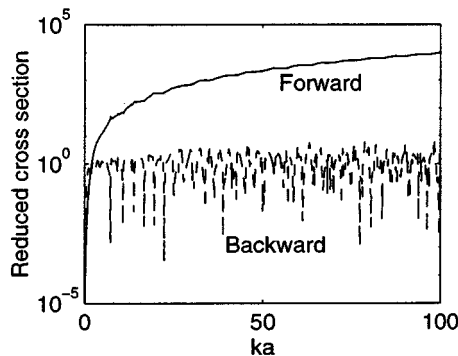


FIG. 4. Reduced scattering cross section versus ka for a stainless-steel sphere.

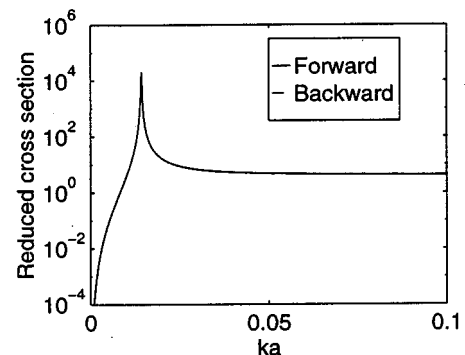


FIG. 5. Reduced scattering cross section versus ka for an air bubble.

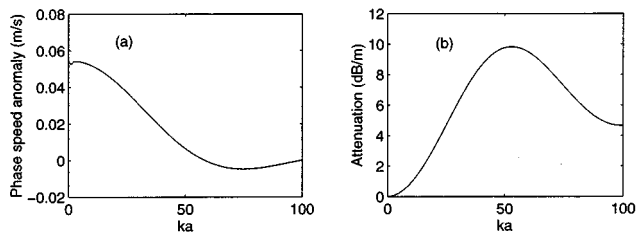


FIG. 6. Phase speed and attenuation versus ka in water containing many low acoustic-contrast fluid spheres for volume fraction $\beta=0.001$.

tively low volume fraction 0.001 the attenuation due to the presence of low acoustic-contrast fluid spheres amounts to as high as about 10 dB/m at $ka=50$, whereas the phase-speed anomaly is less than 6 cm/s throughout the frequencies considered (Fig. 6); the sound-speed anomaly, i.e., Δc , and the attenuation are proportional to the volume fraction in these cases, indicating that the backscattering is negligible, and the dispersion relation of Lax is sufficient. (2) In all cases, as frequency increases, the effective sound speed approaches that in the pure water (1500 m/s), indicating the multiple-scattering effects diminish as the frequency increases. (3) In the low acoustic-contrast fluid sphere case (Fig. 6), the sound-speed anomaly can be either positive or negative, depending on the sign of the real part of scattering function; the attenuation increases with frequency until it reaches a peak at about $ka = 50$. (4) In the rigid fixed sphere case (Fig. 7), interestingly, the phase speed is smaller than that in the pure water, i.e., negative sound-speed anomaly; the sound-speed anomaly, an order of magnitude larger than that of the low acoustic-contrast fluid sphere cases for some frequencies, decreases as the frequency increases. We also note the sound speed dips at low ka [Fig. 7(a) and (c)]; in this case, the attenuation increases smoothly with frequency. The attenuation is of the same order of magnitude as the low acoustic-contrast fluid sphere case. (5) In the elastic sphere case (Fig. 8), the sound phase-speed anomaly is positive and decreases as the frequency increases; we note the peaks in the phase-speed anomaly at small ka , the attenuation generally increases with frequency, with appearance of the oscillatory features at large ka . (6) In the bubbly water case (Fig. 9), the results are consistent with the previous work.¹⁸ Compared to the other cases, the air bubbles have much more significant effects on sound propagation, even at the volume fraction of only 0.001. The phase speed can be increased or decreased by an order of magnitude when the air bubbles are present, while the attenuation can be as large as 1000 dB/m near the

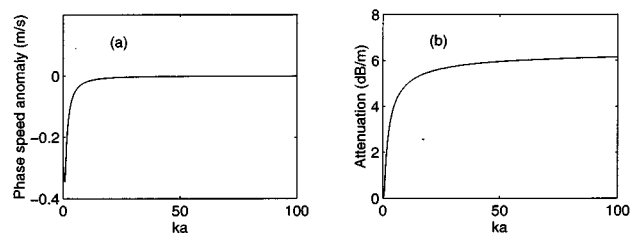


FIG. 7. Phase speed and attenuation versus ka in water containing many rigid fluid spheres for volume fraction $\beta=0.001$.

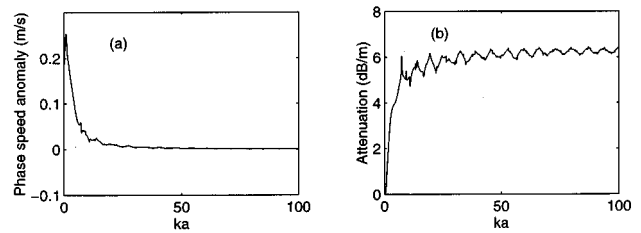


FIG. 8. Phase speed and attenuation versus ka in water containing many stainless-steel spheres for volume fraction $\beta=0.001$.

resonance. A reason why the speed is smaller for the rigid sphere case and bigger for the elastic sphere case may be said as follows. In the rigid sphere case, the acoustic wave cannot penetrate into the scatterer, and must therefore propagate around the scatterer, elongating the travel time. In the elastic sphere case, a portion of the wave can propagate through the scatterer at a greater speed, resulting in an overall reduction in the travel time.

We compare the results shown in Figs. 6–9 with the results from the time-average equation, which was originally derived for the effective velocity as a function of the porosity of the medium,¹⁹

$$\frac{1}{V_e} = \frac{\phi}{V_f} + \frac{1-\phi}{V_m}, \quad (21)$$

where ϕ is the porosity, V_e is the effective velocity, V_f and V_m are velocities of fluid and frame material, respectively. In the present case, the porosity, frame material, and the fluid in the porous region are replaced by the volume fraction, water, and the scatterer, respectively. Therefore we have

$$\frac{1}{v} = \frac{\beta}{c_p} + \frac{1-\beta}{c_w}. \quad (22)$$

Using the parameters given in Table I, the effective sound phase speeds at volume fraction of 0.001 are 1500.057, 1501.17, 1501.50, and 1494.70 for low acoustic-contrast fluid, elastic, rigid, and air-filled spheres, respectively. For the fluid case, Eq. (22) agrees well with the result shown in Fig. 6 as $ka \rightarrow 0$. For the elastic case, the difference of sound-speed anomaly between the two methods is about six times. For the rigid case, the sign of the sound-speed anomaly is different, whereas for the bubble case, not only the sign is different for frequencies above the resonance, but the numerical difference is as large as 175 times for $ka \rightarrow 0$. These results imply that Eq. (22) may agree with Eq. (3) well only

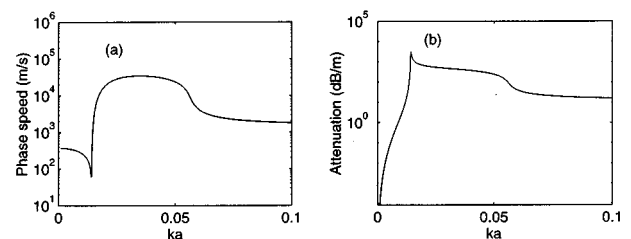


FIG. 9. Phase speed and attenuation versus ka in water containing many weak air bubbles for void fraction $\beta=0.001$.

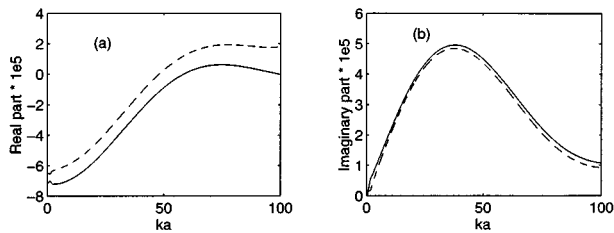


FIG. 10. Comparison of the results from the dispersion relation (8) (solid line) and the results inferred from the KK relations (dashed line) for the low acoustic-contrast fluid sphere case. $\beta=0.001$.

when the sound can easily penetrate the scatterer. In such cases, the scattering is weak.

Next we examine the dispersion via the Kramers–Kronig relations. For convenience, we write the dispersion relation as

$$n_r^2 - 1 \equiv R + iI, \quad (23)$$

where R and I are two real functions. According to the KK relations in Eqs. (4) and (5), the real part of $n_r^2 - 1$ can be converted from the imaginary part, and vice versa. We denote the real and imaginary parts converted from the KK relations as R_{kk} and I_{kk} . Clearly,

$$R_{kk}(\omega) = \frac{2}{\pi} \mathcal{P} \int_0^\infty ds \frac{sI(s)}{s^2 - \omega^2}, \quad (24)$$

and

$$I_{kk}(\omega) = -\frac{2\omega}{\pi} \mathcal{P} \int_0^\infty ds \frac{R(s)}{s^2 - \omega^2}. \quad (25)$$

For the present study, n_r^2 is given by Eq. (8), from which R and I can be deduced. If Eq. (8) describes well the acoustic dispersion in the media considered, it would be expected that R_{kk} and I_{kk} from the KK relations should match R and I , respectively, i.e., R and I satisfy the KK relations. The mismatch between R and R_{kk} , and between I and I_{kk} would indicate the degree of violating the KK relations.

Figures 10–13 compare the real and imaginary parts of $n_r^2 - 1$ to that inferred from the KK relations. In these figures, the solid lines are from Eq. (8) while the dashed lines refer to the results inferred from the KK relations. These figures show the following. (1) The results inferred from the KK relations match the general trend of the results computed directly from the expression in Eq. (8), providing support for the dispersion relation. (2) In the low acoustic contrast

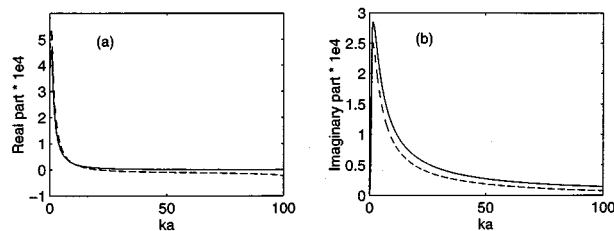


FIG. 11. Comparison of the results from the dispersion relation (8) (solid line) and the results inferred from the KK relations (dashed line) for the rigid fluid case. $\beta=0.001$.

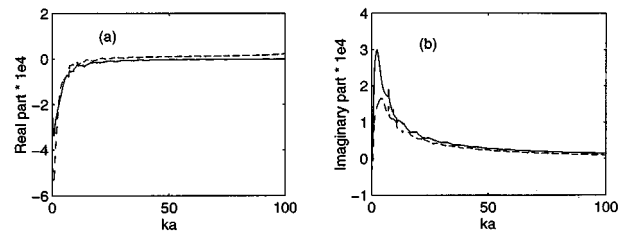


FIG. 12. Comparison of the results from the dispersion relation (8) (solid line) and the results inferred from the KK relations (dashed line) for the elastic sphere case. $\beta=0.001$.

fluid sphere case (Fig. 10), the inferred imaginary part matches the direct result from Eq. (8) rather well, whereas the inferred real part is shifted upward by a nearly constant amount. (3) In the rigid sphere case (Fig. 11), it is the real parts that match each other well, while there is some difference between the imaginary parts; relatively speaking, the difference is more significant when $ka > 2$. (4) In the elastic sphere case (Fig. 12), the feature is similar to that in the rigid sphere case; the difference between the results inferred from the KK relations and that from Eq. (8) is more severe for small ka ; we note the appearance of the dip and peak in the real and imaginary parts, respectively. (5) There appear fine structures in the elastic sphere case, and these fine structures are also inferred from the KK relations (6). For the air bubbles, the results inferred from the KK relations matched the results from the dispersion relation (8) remarkably well (the solid and dashed lines are almost overlapping), indicating that the dispersion relation describes particularly well the acoustic dispersion in bubbly liquids. This could partially explain why relatively more attention has been paid to the acoustic dispersion in bubbly liquids in the past. A further study²⁰ shows that when the thermal and viscosity effects are included, the Foldy's expression for the dispersion relation seems more suitable for larger bubbles than for smaller bubbles; therefore the bubble counting technique based on this dispersion relation may not count small bubbles as accurately as counting larger bubbles. This could partially explain the controversial estimates of small bubbles from different techniques. (7) Speaking overall, the dispersion relation in Eq. (8) best describes the bubble case, the next is the rigid sphere, elastic sphere, and then the low acoustic-contrast fluid sphere.

In the above we have seen that although the results inferred from the KK relations match the general shape of the results from the dispersion relation in Eq. (8), some discrep-

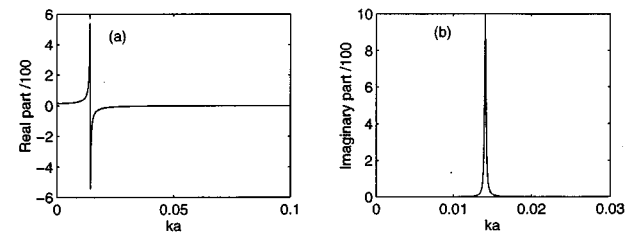


FIG. 13. Comparison of the results from the dispersion relation (8) (solid line) and the results inferred from the KK relations (dashed line) for the air-bubble case. $\beta=0.001$.

ancies remain. Such discrepancies may be related to various assumptions that underlie the dispersion relation. One of the assumptions is that the properties of the individual scatterer are unmodified by the fact that it is embedded in a many body system. As a result, the interaction between particles is represented by the interaction between individual particle and an effective medium. Another assumption is that the scatterers move sufficiently slowly in response to the field acting on them so that their positions can be viewed as adiabatic parameters. Clearly, this assumption may not work well for the low acoustic-contrast fluid sphere; therefore we see larger discrepancies in this case (Fig. 10). For the rigid and stainless-steel spheres, such movement effects are expected to be negligible, which may explain why the dispersion relation in Eq. (8) satisfies the KK relations better than in the low acoustic contrast fluid sphere case. In the bubble case, the movement effects may be surpassed by the dominant resonant scattering, so not to degrade the accuracy of the dispersion relation.

III. SUMMARY

In this paper, we studied acoustic dispersion in various media. The existing dispersion relations have been used to investigate the dispersion in media containing low acoustic-contrast fluid, rigid and elastic spheres, as well as air bubbles. It is shown that in the case of the low acoustic-contrast fluid, rigid and elastic spheres, the presence of scatterers has more profound effects on the acoustic attenuation than on the phase speed, whereas in the bubble case, both the attenuation and phase speed can be affected significantly. The dispersion relations have been further examined using the genuine Kramers–Kronig relations. The results indicate that the existing dispersion relations satisfy the Kramers–Kronig relations generally well. The present work provides a support to the existing dispersion relations.

ACKNOWLEDGMENTS

The author thanks Professor M. S. Wang, Dr. D. Chu, and Dr. M. Treverrow for useful correspondence. Dr. D. Chu

is also thanked for bringing Ref. 19 to the author's attention. The referee is thanked for constructive comments and suggestions, which help improve the presentation of the paper. The work received support from the Department of Fisheries and Oceans, Canada.

- ¹H. Medwin, "In situ acoustic measurements of microbubbles at sea," *J. Geophys. Res.* **82**, 971–976 (1977); N. Breitz and H. Medwin, "Instrumentation for *in situ* acoustical measurements of bubble spectra under breaking waves," *J. Acoust. Soc. Am.* **86**, 739–743 (1989).
- ²G. A. Davidson, "Sound propagation in fogs," *J. Atmos. Sci.* **32**, 2201–2205 (1975).
- ³Z. Ye, T. Curran, and D. Lemon, "Fish detection by the acoustic scintillation technique," *ICES J. Mar. Sci.* **53**, 317–321 (1996).
- ⁴P. D. Thorne, "Measurements of the form function and total scattering cross section for a suspension of spheres," *J. Acoust. Soc. Am.* **93**, 243–248 (1993).
- ⁵A. R. Wenzel and J. B. Keller, "Propagation of acoustic waves in a turbulent medium," *J. Acoust. Soc. Am.* **50**, 911–920 (1971).
- ⁶L. L. Foldy, "The multiple scattering of waves," *Phys. Rev.* **67**, 107–119 (1945).
- ⁷M. Lax, "Multiple scattering of waves," *Rev. Mod. Phys.* **23**, 287–310 (1951).
- ⁸P. C. Waterman and R. Truell, "Multiple scattering of waves," *J. Math. Phys. (N.Y.)* **2**, 512–537 (1961).
- ⁹V. Twersky, "On scattering of waves by random distributors," *J. Math. Phys. (N.Y.)* **3**, 700–715 (1962).
- ¹⁰E. Silberman, "Sound velocity and attenuation in bubbly mixtures measured in standing wave tubes," *J. Acoust. Soc. Am.* **29**, 925–933 (1957).
- ¹¹R. Kronig, "On the theory of dispersion of X-rays," *J. Opt. Soc. Am.* **12**, 547 (1926); R. Kronig and H. A. Kramers, "Absorption and dispersion relations in X-ray spectra," *Z. Phys.* **48**, 174 (1928).
- ¹²C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1995).
- ¹³G. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1985), pp. 421–426.
- ¹⁴Z. Ye and L. Ding, "A study of multiple scattering in bubbly liquids by many body theory," *Can. J. Phys.* **74**, 92–96 (1996).
- ¹⁵V. C. Anderson, "Sound scattering from a fluid sphere," *J. Acoust. Soc. Am.* **22**, 426–431 (1950).
- ¹⁶J. J. Faran, "Sound scattering by solid cylinders and spheres," *J. Acoust. Soc. Am.* **23**, 405–418 (1951).
- ¹⁷R. Hickling, "Analysis of echoes from a solid elastic sphere in water," *J. Acoust. Soc. Am.* **34**, 1582–1592 (1962).
- ¹⁸C. S. Clay and H. Medwin, *Acoustical Oceanography* (Wiley, New York, 1977).
- ¹⁹W. M. Telford, L. P. Geldart, R. E. Sheriff, and D. A. Keys, *Applied Geophysics* (Cambridge U.P., New York, 1976).
- ²⁰Z. Ye, "Kramers–Kronig relations for acoustic dispersion in bubbly liquids" (unpublished).

Reduction of structure-borne noise using an air-voided elastomer^{a)}

Sung H. Ko

Naval Undersea Warfare Center Division, Newport, Rhode Island 02841

(Received 21 July 1996; revised 10 December 1996; accepted 19 December 1996)

A theoretical model was developed to evaluate the reduction of structure-borne noise that is generated by a line force applied on an infinite plate using an air-voided elastomeric baffle. The vibrating plate is covered with an elastomeric baffle layer to reduce the noise generated by the structural vibration. The vibrating plate is perfectly bonded to the elastomeric baffle layer. The outer surfaces of the vibrating plate and the elastomeric baffle are in contact with air and water, respectively. The analysis for modeling is based on the theory of elasticity and pertinent boundary conditions. Effects of various parameters such as baffle layer dimensions and material properties on the noise reduction are presented. [S0001-4966(97)02705-7]

PACS numbers: 43.20.Jr, 43.20.Gp, 43.50.Gf [JEG]

INTRODUCTION

The purpose of this work is to develop a model for evaluating the effectiveness of an air-voided elastomer layer in reducing the structure-borne noise associated with flexural waves.

A simple model for evaluating the flexural wave reduction has been studied using a fluid baffle by Ko *et al.*¹ The model developed in their study is a two-layer plane structure that consists of a vibrating plate and a fluid baffle layer. The analysis has been made using the theory of a thin plate. Further studies were made for three-layer structures using a fluid baffle layer.²⁻⁴ The classical thin plate theory does not account for inertia and shear effects. These effects become important at high frequencies (near the coincidence frequency for a thin plate). In the present study, the theory of elasticity is used in lieu of the thin plate theory. The two-layer structure is comprised of two elastic layers. The outer surfaces of the vibrating plate and the elastomer baffle are in contact with air and water, respectively. The air-voided elastomer layer is perfectly bonded to the vibrating elastic plate. As a related problem, studies on the reflection of a plane acoustic wave incident upon a bilaminar structure and the transmission through the structure have been made by Strifor and Gaunard.⁵

The major results presented in this study are the numerically calculated noise reductions for various parameters.

I. THEORETICAL ANALYSIS

The theoretical model considered for the structure-borne noise reduction using a pressure (point) sensor is depicted in Fig. 1. The upper plate is an elastic plate which is excited by a line force, and the lower plate is an air-voided elastomer layer, which is designed for the reduction of the structure-borne noise.

The governing wave equation in the upper half-space is

$$\frac{\partial^2 p_0}{\partial x^2} + \frac{\partial^2 p_0}{\partial z^2} = \frac{1}{c_0^2} \frac{\partial^2 p_0}{\partial t^2}, \quad (1)$$

where $p_0 = p_0(x, z, t)$ is the acoustic pressure, c_0 is the sound speed in the air, t is the time, and x and z are the spatial coordinates. The acoustic pressure p_0 is written as

$$p_0(x, z, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P_0(k, \omega, z) \times \exp[-i(kx - \omega t)] dk d\omega, \quad (2)$$

where $P_0(k, \omega, z)$ is the Fourier transform of the pressure field $p_0(x, z, t)$, ω is the angular frequency, and k is the wave number in the axial direction. Substituting Eq. (2) into Eq. (1) gives

$$\frac{d^2 P_0}{dz^2} - \alpha_0^2 P_0 = 0, \quad (3)$$

where $\alpha_0 = (k^2 - k_0^2)^{1/2}$ and $k_0 = \omega/c_0$ is the acoustic wave number. The solution of Eq. (3) is written as

$$P_0(k, \omega, z) = A_0(k, \omega) \exp(-\alpha_0 z) \quad (z \geq h_1), \quad (4)$$

where $A_0(k, \omega)$ is the unknown coefficient to be determined. Equation (4) represents the wave pressure that decays away from the plate surface in its normal direction for $k > k_0$ (subsonic wave). However, for the case of $k < k_0$, Eq. (4) represents the wave pressure that propagates in the same direction without decaying (supersonic wave). Therefore, Eq. (4) represents both subsonic and supersonic waves without losing generality.

Similarly, the excitation force on the plate surface ($z = h_1$) in the upper half-space is written as

$$f(x, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(k, \omega) \exp[-i(kx - \omega t)] dk d\omega, \quad (5)$$

where $F(k, \omega)$ is the Fourier transform of the excitation force $f(x, t)$.

^{a)}Paper presented at the 130th Meeting of the Acoustical Society of America, 27 November–1 December 1995, St. Louis, Missouri [J. Acoust. Soc. Am. **98**, 2917(A) (1995)].

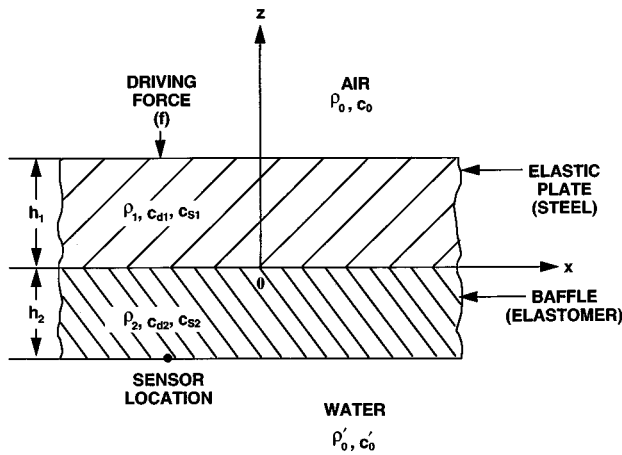


FIG. 1. Geometry of theoretical model.

In the elastic media (both elastic plate and baffle layer), propagation of elastic waves is governed by⁶⁻⁸

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{1}{c_d^2} \frac{\partial^2 \phi}{\partial t^2} \quad (6)$$

and

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{c_s^2} \frac{\partial^2 \psi}{\partial t^2}, \quad (7)$$

where ϕ is the scalar potential, ψ is the y component of the vector potential, and c_d and c_s are the complex dilatational (compressional) and shear (transverse) wave speeds, respectively, and are given by

$$c_d = [(\lambda + 2\mu)/\rho]^{1/2} \quad (8)$$

and

$$c_s = (\mu/\rho)^{1/2}, \quad (9)$$

where ρ is the material density, and λ and μ are the Lamé constants. The scalar and vector component potentials are written as

$$\begin{aligned} \phi(x, z, t) = & \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(k, \omega, z) \\ & \times \exp[-i(kx - \omega t)] dk d\omega \end{aligned} \quad (10)$$

and

$$\begin{aligned} \psi(x, z, t) = & \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Psi(k, \omega, z) \\ & \times \exp[-i(kx - \omega t)] dk d\omega. \end{aligned} \quad (11)$$

Substituting Eqs. (10) and (11) into Eqs. (6) and (7), respectively, one obtains

$$\frac{d^2 \Phi}{dz^2} + (k_d^2 - k^2) \Phi = 0 \quad (12)$$

and

$$\frac{d^2 \Psi}{dz^2} + (k_s^2 - k^2) \Psi = 0, \quad (13)$$

where $k_d = \omega/c_d$ and $k_s = \omega/c_s$ are the dilatational and shear wave numbers, respectively.

The solutions of Eqs. (12) and (13) are written as

$$\Phi(k, \omega, z) = A_1(k, \omega) \cos(\alpha z) + B_1(k, \omega) \sin(\alpha z) \quad (14)$$

and

$$\Psi(k, \omega, z) = A_2(k, \omega) \cos(\beta z) + B_2(k, \omega) \sin(\beta z), \quad (15)$$

where $\alpha = (k_d^2 - k^2)^{1/2}$ and $\beta = (k_s^2 - k^2)^{1/2}$. The normal and tangential displacements are given by

$$u_z = \frac{\partial \phi}{\partial z} + \frac{\partial \psi}{\partial x} \quad (16)$$

and

$$u_x = \frac{\partial \phi}{\partial x} - \frac{\partial \psi}{\partial z}. \quad (17)$$

The normal and tangential stresses are written as

$$\tau_{zz} = (\lambda + 2\mu) \frac{\partial u_z}{\partial z} + \lambda \frac{\partial u_x}{\partial x} \quad (18)$$

and

$$\tau_{zx} = \mu \left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x} \right). \quad (19)$$

The governing wave equation in the lower half-space is

$$\frac{\partial^2 p'_0}{\partial x^2} + \frac{\partial^2 p'_0}{\partial z^2} = \frac{1}{c_0'^2} \frac{\partial^2 p'_0}{\partial t^2}, \quad (20)$$

where $p'_0 = p'_0(x, z, t)$ is the acoustic pressure and c'_0 is the sound speed in the water. The acoustic pressure p'_0 is written as

$$\begin{aligned} p'_0(x, z, t) = & \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P'_0(k, \omega, z) \\ & \times \exp[-i(kx - \omega t)] dk d\omega, \end{aligned} \quad (21)$$

where $P'_0(k, \omega, z)$ is the Fourier transform of the pressure field $p'_0(x, z, t)$. Substituting Eq. (21) into Eq. (20) gives

$$\frac{d^2 P'_0}{dz^2} - \alpha_0'^2 P'_0 = 0, \quad (22)$$

where $\alpha_0' = (k^2 - k_0'^2)^{1/2}$ and $k_0' = \omega/c_0'$ is the acoustic wave number in the water. The solution of Eq. (22) is written as

$$P'_0(k, \omega, z) = A'_0(k, \omega) \exp(\alpha_0' z) \quad (z \leq -h_2), \quad (23)$$

where $A'_0(k, \omega)$ is the unknown coefficient to be determined.

The boundary conditions to be satisfied at the interface between the elastic plate and the air space are that: (1) the pressure acting in the normal direction must be continuous; and (2) the shear stress on the fluid surface is zero. In the normal direction, one writes

$$(\tau_{zz}^{(1)})_{z=h_1} = -(p_0 + f)_{z=h_1}, \quad (24)$$

where $(\tau_{zz}^{(1)})_{z=h_1}$ is the normal stress on the elastic plate surface at $z = h_1$, h_1 is the plate thickness, and the superscript (1) refers to the elastic plate. Further, the acoustic pres-

sure field is coupled to the displacement of the elastic plate by

$$\left[\frac{\partial^2 u_z^{(1)}}{\partial t^2} \right]_{z=h_1} = -\frac{1}{\rho_0} \left[\frac{\partial p_0}{\partial z} \right]_{z=h_1}, \quad (25)$$

where $u_z^{(1)}$ is the normal displacement of the elastic plate. Using Eqs. (2) and (25), one obtains

$$(p_0)_{z=h_1} = -\left[\frac{\rho_0 \omega^2}{\alpha_0} \right] [u_z^{(1)}]_{z=h_1}. \quad (26)$$

Substituting Eq. (26) into Eq. (24), one obtains

$$(\tau_{zz}^{(1)})_{z=h_1} \alpha_0 - \rho_0 \omega^2 (u_z^{(1)})_{z=h_1} = -\alpha_0 f, \quad (27)$$

and the boundary condition for the shear stress at the interface is written as

$$(\tau_{zx}^{(1)})_{z=h_1} = 0. \quad (28)$$

The boundary conditions to be satisfied at the interface between the elastic plate and the baffle layer are that: (1) the normal and tangential stresses; and (2) the normal and tangential displacements must be continuous, i.e.,

$$(\tau_{zx}^{(1)})_{z=0} = (\tau_{zx}^{(2)})_{z=0}, \quad (29)$$

$$(\tau_{zz}^{(1)})_{z=0} = (\tau_{zz}^{(2)})_{z=0}, \quad (30)$$

$$(u_x^{(1)})_{z=0} = (u_x^{(2)})_{z=0}, \quad (31)$$

$$(u_z^{(1)})_{z=0} = (u_z^{(2)})_{z=0}. \quad (32)$$

Note that the superscript (2) refers to the baffle layer.

The boundary conditions to be satisfied at the interface between the baffle layer and the water space are that: (1) the pressure acting in the normal direction must be continuous; and (2) the shear stress on the fluid surface is zero. In the normal direction, one writes

$$(\tau_{zz}^{(2)})_{z=-h_2} = -(p'_0)_{z=-h_2}, \quad (33)$$

where $(\tau_{zz}^{(2)})_{z=-h_2}$ is the normal stress on the baffle surface at $z = -h_2$. Further, the acoustic pressure field is coupled to the displacement of the baffle layer by

$$\left[\frac{\partial^2 u_z^{(2)}}{\partial t^2} \right]_{z=-h_2} = -\frac{1}{\rho'_0} \left[\frac{\partial p'_0}{\partial z} \right]_{z=-h_2}, \quad (34)$$

where $u_z^{(2)}$ is the normal displacement of the baffle layer and ρ'_0 is the water density. Using Eqs. (21) and (34), one obtains

$$(p'_0)_{z=-h_2} = + \left[\frac{\rho'_0 \omega^2}{\alpha'_0} \right] [u_z^{(2)}]_{z=-h_2}. \quad (35)$$

Substituting Eq. (35) into Eq. (33), one obtains

$$(\tau_{zz}^{(2)})_{z=-h_2} \alpha'_0 + \rho'_0 \omega^2 (u_z^{(2)})_{z=-h_2} = 0, \quad (36)$$

and the boundary condition for the shear stress at the interface is written as

$$(\tau_{zx}^{(2)})_{z=-h_2} = 0. \quad (37)$$

Substituting Eqs. (16)–(19) into Eqs. (27)–(32), (36), and (37), one obtains a system of linear algebraic equations

to be solved for undetermined constants $A_1^{(1)}, B_1^{(1)}, \dots, A_2^{(2)}$, and $B_2^{(2)}$ in terms of F_0 as follows:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 & 0 & 0 \\ 0 & a_{32} & a_{33} & 0 & 0 & a_{36} & a_{37} & 0 \\ a_{41} & 0 & 0 & a_{44} & a_{45} & 0 & 0 & a_{48} \\ a_{51} & 0 & 0 & a_{54} & a_{55} & 0 & 0 & a_{58} \\ 0 & a_{62} & a_{63} & 0 & 0 & a_{66} & a_{67} & 0 \\ 0 & 0 & 0 & 0 & a_{75} & a_{76} & a_{77} & a_{78} \\ 0 & 0 & 0 & 0 & a_{85} & a_{86} & a_{87} & a_{88} \end{bmatrix} \begin{bmatrix} A_1^{(1)} \\ B_1^{(1)} \\ A_2^{(1)} \\ B_2^{(1)} \\ A_1^{(2)} \\ B_1^{(2)} \\ A_2^{(2)} \\ B_2^{(2)} \end{bmatrix} = \begin{bmatrix} b_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (38)$$

where $b_1 = -F_0 \alpha_0$ and a_{11}, a_{12}, \dots , and a_{88} are shown in the Appendix. If a harmonic line force, $f(x, t) = F_0 \delta(x) \exp(-i\omega' t)$, is acting on the elastic plate (air side), then its Fourier transform is given by $F(k, \omega) = F_0 \delta(\omega - \omega')$, where F_0 is a constant force per unit area, and $\delta(x)$ and $\delta(\omega - \omega')$ are the delta functions. If Eq. (23) is substituted into Eq. (21), then the pressure field at the sensor location (baffle surface in the water) is written as

$$\begin{aligned} [p'_0(z, \omega)]_{z=-h_2} &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} [A'_0(k, \omega, z) \exp(\alpha'_0 z)]_{z=-h_2} dk \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\rho'_0 \omega^2}{F_0 \alpha'_0} [U_z^{(2)}(k, \omega, z)]_{z=-h_2} dk. \end{aligned} \quad (39)$$

Note that the Fourier transform of the normal displacement of the baffle layer is written as

$$\begin{aligned} [U_z^{(2)}(k, \omega, z)]_{z=-h_2} &= [-\alpha_2 \sin(\alpha_2 z) A_1^{(2)} + \alpha_2 \cos(\alpha_2 z) B_1^{(2)} \\ &\quad - ik \cos(\beta_2 z) A_2^{(2)} - ik \sin(\beta_2 z) B_2^{(2)}]_{z=-h_2}, \end{aligned} \quad (40)$$

where α_2 and β_2 are the values of α and β defined in Eqs. (14) and (15) for the baffle layer. It should be noted that the constants $A_1^{(2)}, B_1^{(2)}, A_2^{(2)}$, and $B_2^{(2)}$ are expressed in terms of F_0 . In the absence of the baffle layer, the pressure field at the sensor location (elastic plate surface in the water) is written as

$$[p'_0(z, \omega)]_{z=0} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [A'_0(k, \omega, z) \exp(\alpha'_0 z)]_{z=0} dk$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\rho'_0 \omega^2}{F_0 \alpha'_0} [U_z^{(1)}(k, \omega, z)]_{z=0} dk. \quad (41)$$

It is important to mention that the case of no baffle layer requires the analysis of the elastic plate only to obtain the Fourier transform of the normal displacement as shown in Eq. (41). Using the pertinent boundary conditions for this analysis, one obtains the following system of linear algebraic equations to be solved for the undetermined coefficients $A_1^{(1)}$, $B_1^{(1)}$, $A_2^{(1)}$, and $B_2^{(1)}$ in terms of F_0 :

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} A_1^{(1)} \\ B_1^{(1)} \\ A_2^{(1)} \\ B_2^{(1)} \end{bmatrix} = \begin{bmatrix} b_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (42)$$

where a_{11}, a_{12}, \dots , and a_{24} are the same coefficients as those used in Eq. (38). Other coefficients are

$$a_{32} = -2i\rho_1 c_{s1}^2 k \alpha_1, \quad a_{33} = -\rho_1 c_{s1}^2 (k^2 - \beta_1^2),$$

$$a_{41} = +\rho_1 c_{s1}^2 (k^2 - \beta_1^2) \alpha'_0, \quad a_{42} = +\rho'_0 \omega^2 \alpha_1, \quad (43)$$

$$a_{43} = -ik\rho'_0 \omega^2, \quad a_{44} = -2i\rho_1 c_{s1}^2 k \beta_1 \alpha'_0,$$

where α_1 and β_1 are the values of α and β defined in Eqs. (14) and (15) for the elastic plate, and c_{s1} is the complex shear wave speed in the elastic plate. Then, the Fourier transform of the normal displacement is obtained as

$$[U_z^{(1)}(k, \omega, z)]_{z=0} = [-\alpha_1 \sin(\alpha_1 z) A_1^{(1)} + \alpha_1 \cos(\alpha_1 z) B_1^{(1)} - ik \cos(\beta_1 z) A_2^{(1)} - ik \sin(\beta_1 z) B_2^{(1)}]_{z=0}. \quad (44)$$

The frequency spectral density for a point sensor represents the response of the sensor to the excitation force acting on the air side. In the presence of the baffle layer, it is written as

$$Q_b(\omega) = 2\pi \int_{-\infty}^{+\infty} T_b(k, \omega) dk, \quad (45)$$

where $T_b(k, \omega)$ is the transfer function in the presence of the baffle layer and is given by

$$T_b(k, \omega) = \left| \frac{\rho'_0 \omega^2}{F_0 \alpha'_0} [U_z^{(2)}(k, \omega, z)]_{z=-h_2} \right|^2. \quad (46)$$

In the absence of the baffle layer, it is written as

$$Q_0(\omega) = 2\pi \int_{-\infty}^{+\infty} T_0(k, \omega) dk, \quad (47)$$

where $T_0(k, \omega)$ is the transfer function in the absence of the baffle layer and is given by

$$T_0(k, \omega) = \left| \frac{\rho'_0 \omega^2}{F_0 \alpha'_0} [U_z^{(1)}(k, \omega, z)]_{z=0} \right|^2. \quad (48)$$

The structure-borne noise reduction is defined as

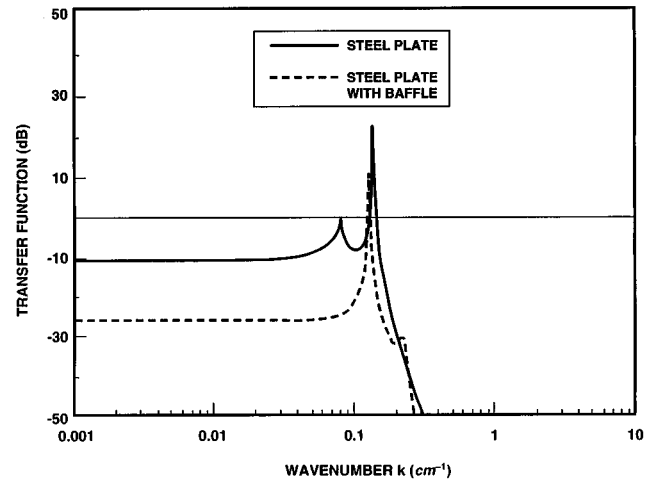


FIG. 2. Transfer functions calculated at 2000 Hz for the 5.08-cm (2-in.) vibrating plate with and without a 7.62-cm (3-in.) elastomer baffle.

$$NR = 10 \log_{10} \left[\frac{Q_b(\omega)}{Q_0(\omega)} \right] \quad (\text{dB}), \quad (49)$$

where $Q_b(\omega)$ and $Q_0(\omega)$ are given by Eqs. (45) and (47).

II. NUMERICAL RESULTS AND DISCUSSION

The flexural wave noise levels with and without the baffle can be calculated by evaluating frequency spectral densities given by Eqs. (45) and (47), respectively. It is important to analyze integrands, i.e., transfer functions $T_b(k, \omega)$ and $T_0(k, \omega)$, to understand frequency wave-number spectra. The baseline data used in the numerical calculations are as follows:

| | |
|---|----------------------------|
| ρ_0 (air density) | 0.001 21 g/cm ³ |
| c_0 (sound speed in air) | 34 000 cm/s |
| ρ'_0 (water density) | 1.0 g/cm ³ |
| c'_0 (sound speed in water) | 150 000 cm/s |
| h_1 (steel plate thickness) | 5.08 cm (2 in.) |
| ρ_1 (steel density) | 7.8 g/cm ³ |
| c_{s01} (shear wave speed in steel) | 310 000 cm/s |
| c_{d01} (dilatational wave speed in steel) | 579 000 cm/s |
| ζ_{s1} (shear loss factor of steel) | 0.005 |
| ζ_{d1} (dilatational loss factor of steel) | 0.001 |
| h_2 (baffle layer thickness) | 7.62 cm (3 in.) |
| ρ_2 (baffle density) | 0.6 g/cm ³ |
| c_{s02} (shear wave speed in baffle) | 5000 cm/s |
| c_{d02} (dilatational wave speed in baffle) | 50 000 cm/s |
| ζ_{s2} (shear loss factor of baffle) | 0.3 |
| ζ_{d2} (dilatational loss factor of baffle) | 0.05. |

The subscripts 1 and 2 in the above baseline data set refer to the elastic plate and baffle layer, respectively. The complex shear and dilatational wave speeds are written, respectively, as

$$c_s = c_{s0}(1 + i\zeta_s)^{1/2} \quad (50)$$

and

$$c_d = c_{d0}(1 + i\zeta_d)^{1/2}, \quad (51)$$

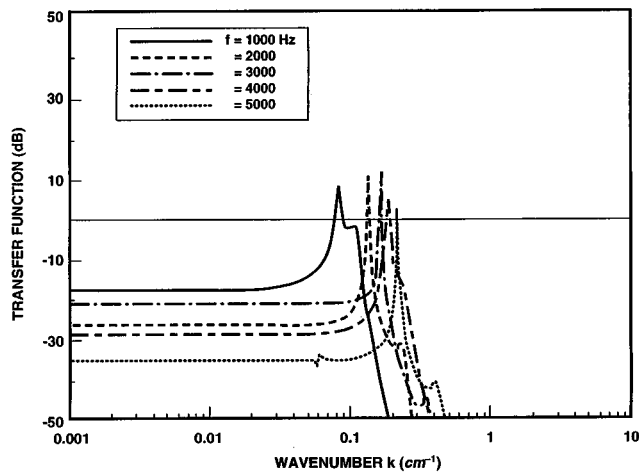


FIG. 3. Transfer functions calculated at various frequencies for the 5.08-cm vibrating plate covered with a 7.62-cm elastomer baffle.

where c_{s0} and c_{d0} are the real parts of the complex shear and dilatational wave speeds, and ζ_s and ζ_d are the loss factors associated with the respective waves.

Figure 2 shows the numerical results calculated for the transfer functions as a function of wave number k at 2000 Hz. The solid and dashed curves are the transfer functions calculated for the 2-in.-thick steel plate [using Eq. (48)] and for the 2-in.-thick steel plate covered with the 3-in.-thick baffle [using Eq. (46)], respectively. Figure 3 presents the transfer functions calculated for the 2-in.-thick steel plate with the 3-in.-thick baffle at various frequencies. As can be seen in the figure, flexural wave-number peaks increase (slightly shift to the right) as the frequency increases.

Figures 4–9 present noise reduction curves as a function of frequency. Figure 4 presents noise reductions for various thicknesses of the baffle layer. The solid, dashed, chain-dotted, and chain-dashed lines denote the noise reduction curves calculated for 1-, 2-, 3-, and 4-in.-baffle layers perfectly bonded to a 2-in.-thick steel plate. As can be seen in the figure, the maximum amount of the noise reduction depends on the resonance phenomena caused by the thickness of a given material. It is noted that a thicker (thinner) layer is

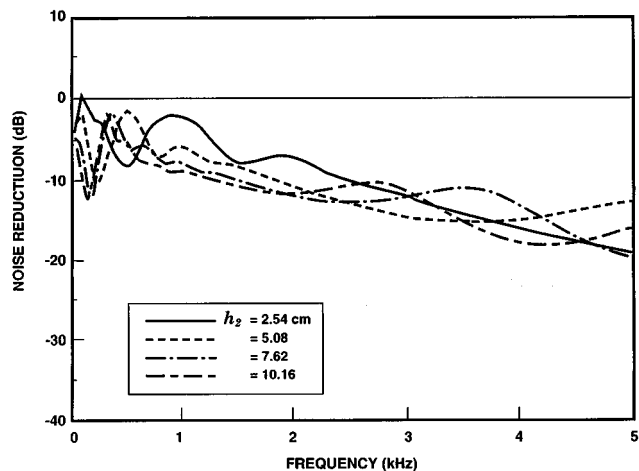


FIG. 4. Noise reduction curves for various baffle layer thicknesses (h_2).

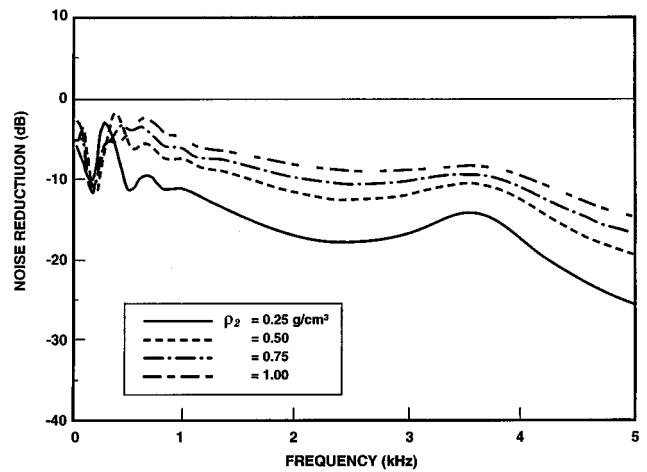


FIG. 5. Noise reduction curves for various baffle layer densities (ρ_2).

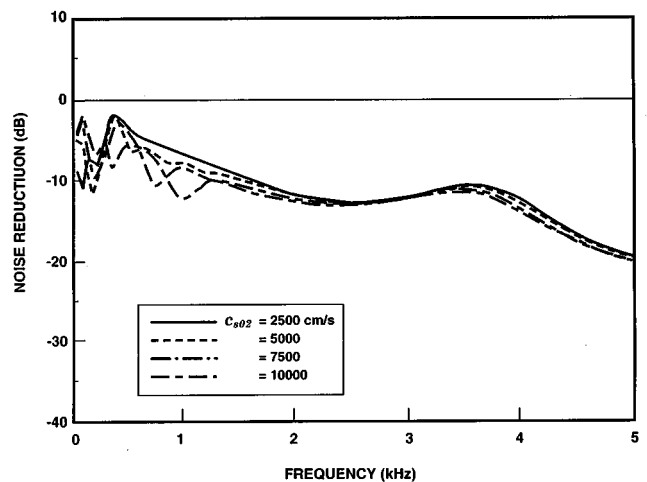


FIG. 6. Noise reduction curves for various shear wave speeds (c_{s02}) in the baffle layer.

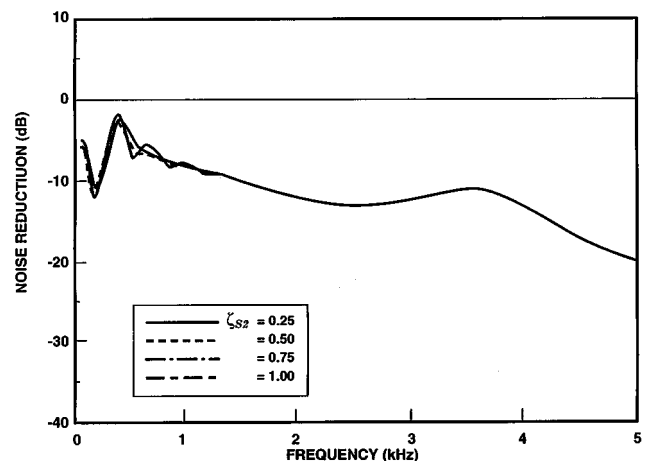


FIG. 7. Noise reduction curves for various loss factors (ζ_{s2}) associated with the shear wave in the baffle layer.

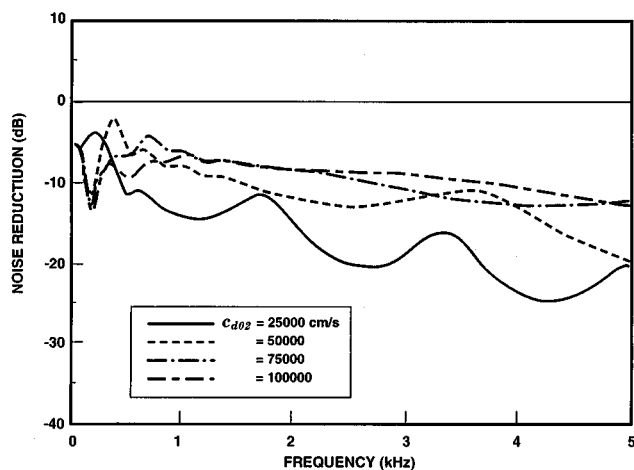


FIG. 8. Noise reduction curves for various dilatational wave speeds (c_{d02}) in the baffle layer.

better for the low (high) frequency. Figure 5 shows the effect of the baffle density (ρ_2) on the noise reduction. It is shown that more noise reduction is achieved as the baffle density becomes small for a given baffle layer thickness, except in the low-frequency range. Since the lighter the baffle is, the lower the density, a lighter baffle is desirable for the noise reduction.

Figure 6 shows the effect of the shear wave speed (c_{s02}) on the noise reduction. As shown in Fig. 6, almost no effect has been observed above 1200 Hz for the assigned values of c_{s02} . However, below 1200 Hz, some effect is noticed depending on the frequency. Figure 7 shows the effect of the loss factor (ζ_{s2}) associated with the shear wave speed. It is shown that no effect on the loss factor is observed for the assigned values of ζ_{s2} for the entire frequency range.

Figure 8 presents the effect of the dilatational wave speed (c_{d02}) in the baffle layer on the noise reduction. As shown in Fig. 8, more noise reduction is achieved as the value of c_{d02} decreases. Note, again, that the lighter the baffle is, the lower the dilatational wave speed. This implies that a light baffle material is more effective than a heavy one. The resonance phenomena caused by the dilatational wave

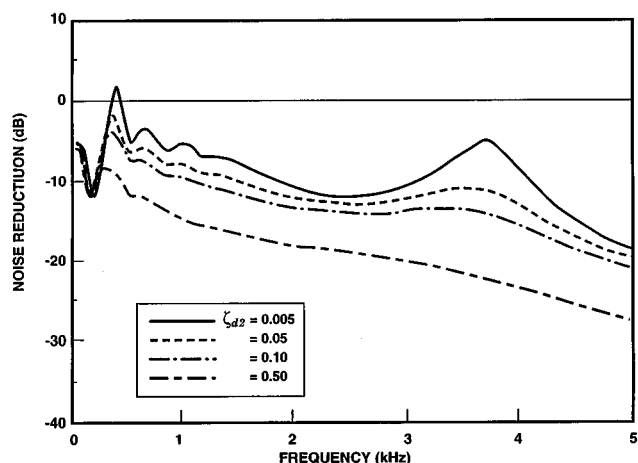


FIG. 9. Noise reduction curves for various loss factors (ζ_{d2}) associated with the dilatational wave in the baffle layer.

speed for a given baffle layer thickness are more clearly shown in the Fig. 8. Figure 9 presents the effect of the loss factor (ζ_{d2}) associated with the dilatational wave in the baffle layer on the noise reduction. As shown in Fig. 9, more noise reduction is achieved as the value of ζ_{d2} increases.

III. CONCLUDING REMARKS

A theoretical model has been developed for reducing the structure-borne noise using an air-voided elastomeric baffle layer. The theoretical model used here is an elastic plate covered with a layer of elastomer that serves as a flexural wave baffle. Based on the limited study conducted in this work, the following conclusions are drawn:

- (1) The amount of noise reduction increases as (a) the baffle layer thickness increases and (b) the loss factor associated with the dilatational wave increases.
- (2) The amount of noise reduction increases as (a) the dilatational wave speed in the baffle decreases and (b) the baffle density decreases. This implies that a lighter material is more suitable for the reduction of the structure-borne noise.
- (3) The effects of the shear wave speed in the baffle and the loss factor associated with the shear wave on the noise reduction are insignificant.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Dr. Howard H. Schloemer for his helpful discussion. This work was supported by the Office of Naval Research, Code 321SS.

APPENDIX

$$\begin{aligned}
 a_{11} &= \rho_1 c_{s1}^2 (k^2 - \beta_1^2) \alpha_0 \cos(\alpha_1 h_1) + \rho_0 \omega^2 \alpha_1 \sin(\alpha_1 h_1), \\
 a_{12} &= \rho_1 c_{s1}^2 (k^2 - \beta_1^2) \alpha_0 \sin(\alpha_1 h_1) - \rho_0 \omega^2 \alpha_1 \cos(\alpha_1 h_1), \\
 a_{13} &= 2i \rho_1 c_{s1}^2 k \beta_1 \alpha_0 \sin(\beta_1 h_1) + i \rho_0 \omega^2 k \cos(\beta_1 h_1), \\
 a_{14} &= -2i \rho_1 c_{s1}^2 k \beta_1 \alpha_0 \cos(\beta_1 h_1) + i \rho_0 \omega^2 k \sin(\beta_1 h_1), \\
 a_{21} &= 2i \rho_1 c_{s1}^2 k \alpha_1 \sin(\alpha_1 h_1), \\
 a_{22} &= -2i \rho_1 c_{s1}^2 k \alpha_1 \cos(\alpha_1 h_1), \\
 a_{23} &= -\rho_1 c_{s1}^2 (k^2 - \beta_1^2) \cos(\beta_1 h_1), \\
 a_{24} &= -\rho_1 c_{s1}^2 (k^2 - \beta_1^2) \sin(\beta_1 h_1), \\
 a_{32} &= -2i \rho_1 c_{s1}^2 k \alpha_1, \quad a_{33} = -\rho_1 c_{s1}^2 (k^2 - \beta_1^2), \\
 a_{36} &= 2i \rho_2 c_{s2}^2 k \alpha_2, \quad a_{37} = \rho_2 c_{s2}^2 (k^2 - \beta_2^2), \\
 a_{41} &= \rho_1 c_{s1}^2 (k^2 - \beta_1^2), \quad a_{44} = -2i \rho_1 c_{s1}^2 k \beta_1, \\
 a_{45} &= -\rho_2 c_{s2}^2 (k^2 - \beta_2^2), \quad a_{48} = 2i \rho_2 c_{s2}^2 k \beta_2, \\
 a_{51} &= -ik, \quad a_{54} = -\beta_1, \quad a_{55} = +ik, \quad a_{58} = +\beta_2, \\
 a_{62} &= +\alpha_1, \quad a_{63} = -ik, \quad a_{66} = -\alpha_2, \quad a_{67} = +ik, \\
 a_{75} &= \rho_2 c_{s2}^2 (k^2 - \beta_2^2) \alpha'_0 \cos(\alpha_2 h_2) + \rho'_0 \omega^2 \alpha_2 \sin(\alpha_2 h_2),
 \end{aligned}$$

$$\begin{aligned}
a_{76} &= -\rho_2 c_{s2}^2 (k^2 - \beta_2^2) \alpha'_0 \sin(\alpha_2 h_2) \\
&\quad + \rho'_0 \omega^2 \alpha_2 \cos(\alpha_2 h_2), \\
a_{77} &= -2i\rho_2 c_{s2}^2 k \beta_2 \alpha'_0 \sin(\beta_2 h_2) - i\rho'_0 \omega^2 k \cos(\beta_2 h_2), \\
a_{78} &= -2i\rho_2 c_{s2}^2 k \beta_2 \alpha'_0 \cos(\beta_2 h_2) + i\rho'_0 \omega^2 k \sin(\beta_2 h_2), \\
a_{85} &= -2i\rho_2 c_{s2}^2 k \alpha_2 \sin(\alpha_2 h_2), \\
a_{86} &= -2i\rho_2 c_{s2}^2 k \alpha_2 \cos(\alpha_2 h_2), \\
a_{87} &= -\rho_2 c_{s2}^2 (k^2 - \beta_2^2) \cos(\beta_2 h_2), \\
a_{88} &= \rho_2 c_{s2}^2 (k^2 - \beta_2^2) \sin(\beta_2 h_2).
\end{aligned} \tag{2}$$

- ¹S. H. Ko, C. H. Sherman, and W. A. Strawderman, "Baffling of flexural waves: Noise reduction," Naval Underwater Systems Center Tech. Memo. No. TD12-80-74, October 1974.
- ²S. H. Ko and C. H. Sherman, "Effectiveness of flexural wave baffles for improving performance of hull mounted arrays," Naval Underwater Systems Center Tech. Memo. No. 771062, April 1977.
- ³S. H. Ko and C. H. Sherman, "Flexural wave baffling," J. Acoust. Soc. Am. **66**, 566-570 (1979).
- ⁴S. H. Ko, "Flexural wave reduction using a compliant tube baffle," J. Acoust. Soc. Am. **99**, 691-699 (1996).
- ⁵H. C. Strifors and G. C. Gaunard, "Selective reflectivity of viscoelastically coated plates in water," J. Acoust. Soc. Am. **88**, 901-910 (1990).
- ⁶H. Kolsky, *Stress Wave in Solids* (Dover, New York, 1963).
- ⁷T. R. Meeker and A. H. Meizler, "Guided wave propagation in elongated cylinders and plates," in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1964).
- ⁸J. D. Achenbach, *Wave Propagation in Elastic Solid* (North-Holland, New York, 1973).

A combined integro-modal approach for predicting acoustic properties of irregular-shaped cavities

J. Missaoui and L. Cheng

Department of Mechanical Engineering, Laval University, Québec, Québec G1K 7P4, Canada

(Received 8 May 1996; accepted for publication 10 February 1997)

An integro-modal approach is presented in this paper for computing the acoustic properties of irregular-shaped cavities. The method consists of discretizing the whole cavity into a series of subcavities, whose acoustic pressure is decomposed either over a modal basis of regular subcavities or over that of the bounding cavities in the case of irregular-shaped boundaries. An integral formulation is then established to ensure continuity of both pressure and velocity between adjacent subcavities using a membrane with zero mass and stiffness. To some extent, the method provides a combined approach retaining the advantages of both the acoustoelastic method (AEM) and the Green's function method (GFM). Numerical and experimental results are presented demonstrating the efficiency and accuracy of the suggested technique. Comparisons with other existing methods are also made. It is shown that good accuracy on the computation of cavity modes can be obtained using a very limited number of subcavities. © 1997 Acoustical Society of America. [S0001-4966(97)04806-6]

PACS numbers: 43.20.Ks, 43.55.Ka [ANN]

INTRODUCTION

Interior noise associated with surrounding vibrating structures has been the subject of many studies. One typical example is the cabin noise of aircraft and vehicles with vibrating walls which frequently create noise levels exceeding the human comfort limits. In such applications, irregular-shaped cavities are often involved, whose modal characteristics are fundamental to a better understanding of the sound radiation mechanism and any noise reduction action that needs to be taken.

Analytical expressions are available to calculate natural frequencies and mode shapes of regular cavities. Unfortunately, only a few simple geometries such as rectangular and cylindrical cavities allowing the separation of the variables can be simply treated by classical methods. As far as other cavities with an arbitrary shape are concerned, the literature shows that the available methods are quite limited. The most popular alternatives are numerical methods¹⁻⁴ such as finite element method (FEM). This method has been widely and successfully used in the past to analyze irregular cavities. The main disadvantage is the large number of degrees required and the related computation time. Other methods involving a more physical basis rather than numerical ones have been also developed. In Ref. 5 an acoustoelastic model (AEM) was used to determine acoustic natural frequencies of multiply connected regular cavities. The method was then extended in Ref. 6 to compute the acoustic modal properties of irregular-shaped cavity. The procedure is based on an approximation of the cavity geometry by a set of rectangular subcavities. Adjacent subcavities are jointed together by means of vibrating membranes. Although only rectangular subcavities were used in this work, it is natural to suppose that other regular subcavities can also be used if they admit analytical solutions. However, the disadvantage of this approach lies in the fact that the number of subcavities required is highly dependent on the irregularity of the cavity shape.

Approximating the cavity geometry by a series of regular shapes introduces inevitable errors in the computation. Another approach, based on the Green's function method (GFM) was described in Ref. 7. A practical application of this method was reported by Succi⁸ who calculated the acoustic response of an automobile cabin by imposing the vibration of panels on the cavity boundary. The method is very helpful since there is no need to calculate the acoustic modes as far as the structural-acoustic response is concerned. The method is limited however to a cavity only slightly distorted from a regular one.

In the present paper an alternative approach is presented for computing the acoustic properties of cavities of arbitrary shape. This method is based on an integro-modal formulation combining the advantages of the two methods mentioned above (AEM and GFM). It permits the use of mixed subcavities, of either regular or irregular shape, leading to a minimal discretization. The modal characteristics of regular subcavities are obtained analytically while the irregular subcavities are treated using a modal expansion over the mode shapes of their regular bounding cavities. An integral formulation is then established to ensure both the continuity of the pressure and velocity between adjacent subcavities using a membrane with zero mass and stiffness. In regard to AEM, both regular and irregular subcavities can be used in this approach. In doing this the approximation is related to the solution, rather than the cavity shape. One would expect that the solution converges more rapidly using fewer subcavities. Compared to GFM, the fact that the bounding cavities can be taken for each subcavities makes it possible to choose envelopes even if the cavity is strongly irregular. As a result, envelopes closer to the subcavity geometries can be used to enhance the prediction accuracy.

The mathematical formulation is described in Sec. I for an irregular-shaped cavity. Numerical results are then presented in Sec. II. Several configurations are then used to

study the convergence of the method and to validate the theory. In each case, resonant frequencies of the cavity computed using the present approach are compared with results reported in the literature. An experimental validation is also reported in Sec. III.

I. FORMULATION

The mathematical formulation consists in treating an irregular-shaped cavity as a combination of connected subcavities separated by elastic panels. In each subcavity, the interior sound pressure p can be calculated by transforming the Helmholtz equation into an integral form via the second Green's theorem.

$$\int_V G(\nabla^2 p + \lambda^2 p) dv = \int_V p(\nabla^2 G + \lambda^2 G) dv + \int_{S_b} \left(G \frac{\partial p}{\partial \mathbf{n}} - p \frac{\partial G}{\partial \mathbf{n}} \right) ds, \quad (1)$$

where λ is the wave number; \mathbf{n} the outward normal vector of the boundary surface S_b of the enclosure with volume V ; G the Green's function corresponding to a transfer function obtained between an observation point (r) and the source (r_0). The construction of the function G for a Neumann boundary is based on the inhomogeneous Helmholtz equation with an infinite surface impedance written as follows:

$$\nabla^2 G(r, r_0) + \lambda^2 G(r, r_0) = -\delta(r - r_0),$$

$$\frac{\partial G(r, r_0)}{\partial \mathbf{n}} = 0, \quad (2)$$

in which $\delta(r - r_0)$ is the Dirac delta function. The analytical expression for the function G satisfying Eq. (2) and using normal modal expansion can be written as follows:

$$G(r, r_0) = \sum_n \frac{c^2 \varphi_n(r) \varphi_n(r_0)}{(\omega_n^2 - \omega^2) V \wedge_n}, \quad (3)$$

where c is the speed of the sound in the internal medium, ω_n is the angular resonance frequency of the cavity, φ_n the corresponding mode shape and $\wedge_n = (1/V) \int_V \varphi_n^2(r) dv$ the generalized acoustic mass.

An irregular-shaped cavity may be composed of both regular and irregular subcavities, each of which being treated differently hereafter. For illustration purposes, the formulation is first developed for the cavity shown in Fig. 1(a). The whole process will then be generalized. As illustrated in Fig. 1, the cavity investigated can be divided into a regular [Fig. 1(b)] and an irregular [Fig. 1(c)] subcavity. The junction between the two subcavities is replaced by a vibrating panel. These two subcavities are first treated separately and then coupled together by harmonizing the motion of the panel.

A. Regular subcavity

In the case of a regular-shaped cavity, analytical expressions are available for the mode shapes and the natural frequencies. Typical examples are rectangular, cylindrical, or semi-cylindrical enclosures. Based on a fluid formulation,

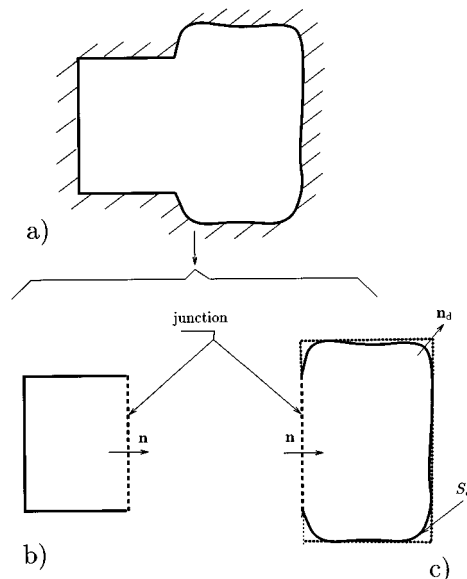


FIG. 1. Discretization procedure used in the present approach. (a) Real cavity; (b) regular subcavity; panel: -----; (c) irregular subcavity; panel: -----; bounding cavity:

the internal pressure is decomposed using an orthogonal expansion in terms of the hard-walled cavity modes:

$$p = \rho_f c^2 \sum_n \frac{a_n(t)}{\wedge_n} \varphi_n(r), \quad (4)$$

where ρ_f is the fluid density, n the modal indices of the cavity, and $a_n(t)$ the modal pressure amplitudes to be determined. The transverse displacement of the vibrating wall w is also expanded in terms of *in vacuo* normal mode shapes ψ_m :

$$w = \sum_m q_m(t) \psi_m, \quad (5)$$

where m contains the structural modal indices; $q_m(t)$ are the structural modal coordinates. Assuming that no absorbent boundary conditions are present and that the interior noise is due to arbitrary vibrating surfaces with the remaining part being acoustically hard wall, the substitution of Eqs. (3), (4), and (5) in Eq. (1) leads to a linear modal acoustic equation:

$$\ddot{a}_n + \omega_n^2 a_n = -\frac{A_f}{V} \sum_m \ddot{q}_m L_{nm}, \quad (6)$$

where A_f is the area of the vibrating surface and

$$L_{nm} = \frac{1}{A_f} \int_s \varphi_n(r) \psi_m(r) ds, \quad (7)$$

with L_{nm} being the modal coupling coefficient between the m th structure mode and the n th cavity mode. This term characterizes the coupling in space between the two modes.

B. Irregular subcavity

An irregular-shaped subcavity may be considered as a deviation from a regular shape, as shown in Fig. 1(c). In this section the procedure consists of enclosing the irregular cavity by a regular one, called the envelope or bounding cavity,

for which modal information is available. Since the natural modes of the irregular-shaped cavity are not known analytically, the modes of the bounding cavity ϕ are then used to perform the pressure decomposition in a similar way to that illustrated in Eq. (3), by replacing φ by ϕ . Similarly, the Green's function can also be obtained. Except for the flexible part, all the remaining boundary are supposed to be acoustically hard. Equation (1) hence becomes

$$\begin{aligned} \sum_n \frac{a_n(t)}{\Lambda_n} \phi_n(r) = & -\frac{c^2}{V} \sum_{n'} \sum_n \frac{\phi_n(r)}{(\omega_n^2 - \omega^2) \Lambda_n} \frac{a_{n'}(t)}{\Lambda_{n'}} \\ & \times \int_{S_d} \phi_{n'}(r_0) \frac{\partial \phi_n(r_0)}{\partial \mathbf{n}_d} ds \\ & - \frac{A_f}{V} \sum_m \sum_n \frac{\ddot{q}_m \phi_n(r) L_{nm}}{\Lambda_n (\omega_n^2 - \omega^2)}, \end{aligned} \quad (8)$$

where V , ω_n , and Λ_n belong to the bounding cavity; S_d is the surface of the irregular subcavity with a unit normal vector \mathbf{n}_d . L_{nm} is again the coupling coefficient defined in Eq. (7), after replacing φ by ϕ . Using the orthogonality principle of the eigenfunction of the bounding cavity, the following modal equation is obtained:

$$\begin{aligned} \ddot{a}_n + \omega_n^2 a_n + \frac{c^2}{V} \sum_{n'} \frac{a_{n'}}{\Lambda_{n'}} \int_{S_d} \phi_{n'} \frac{\partial \phi_n}{\partial \mathbf{n}_d} ds \\ = -\frac{A_f}{V} \sum_m \ddot{q}_m L_{nm}. \end{aligned} \quad (9)$$

In Eq. (9), the integration is performed over the surface of the irregular cavity, either analytically or numerically, depending on the complexity of the shape. One can see that any irregularity of the boundary shape has the effect of coupling the acoustic modes of its envelope. One property of the integral term can be evaluated from the Helmholtz equation by the following simple expression:

$$\begin{aligned} (\omega_n^2 - \omega_{n'}^2) \int_{V_d} \phi_n \phi_{n'} dv = c^2 \left(\int_{S_d} \phi_{n'} \frac{\partial \phi_n}{\partial \mathbf{n}_d} ds \right. \\ \left. - \int_{S_d} \phi_n \frac{\partial \phi_{n'}}{\partial \mathbf{n}_d} ds \right). \end{aligned} \quad (10)$$

In the special case where both cavities coincide, the hard wall condition gives:

$$\int_S \phi_{n'} \frac{\partial \phi_n}{\partial \mathbf{n}} ds = \int_S \phi_n \frac{\partial \phi_{n'}}{\partial \mathbf{n}} ds = 0. \quad (11)$$

Consequently, Eq. (10) shows the orthogonality relations between the acoustic modes

$$\int_V \phi_n \phi_{n'} dv = 0; \quad n \neq n'. \quad (12)$$

It should be observed that Eq. (9) is a more general form than Eq. (6). Hence, in the following development Eq. (9) is retained as a characteristic equation of the subcavities.

C. Coupling between the two subcavities

The cavity illustrated in Fig. 1(a) has been discretized to a regular and an irregular subcavity, connected by elastic panel, as shown in Fig. 1(b) and (c). The acoustic pressure difference across the interconnected region is governed by the vibrations of the elastic panel. The pressure jump \bar{p} and the normal gradient pressure jump on both sides (+ and -) of the panel can be expressed respectively as:

$$\begin{aligned} \bar{p} = \rho_f c^2 \left(\sum_n \frac{a_n}{\Lambda_n} \phi_n \right)^+ - \rho_f c^2 \left(\sum_n \frac{a_n}{\Lambda_n} \phi_n \right)^-, \\ \frac{\partial \bar{p}}{\partial \mathbf{n}} = \rho_f c^2 \left(\sum_n \frac{a_n}{\Lambda_n} \frac{\partial \phi_n}{\partial \mathbf{n}^+} \right)^+ - \rho_f c^2 \left(\sum_n \frac{a_n}{\Lambda_n} \frac{\partial \phi_n}{\partial \mathbf{n}^-} \right)^-. \end{aligned} \quad (13)$$

These two quantities depend on the structural properties. In fact, the panel is exposed to an acoustic pressure loading on both sides. Assuming a harmonic behavior for the whole system [$a_n(t) = P_n \sin(\omega t)$ for sound pressure and $q_m(t) = U_m \sin(\omega t)$ for the panel], the governing equation of motion of a thin isotropic panel may be used:

$$D_s \nabla^4 w + \rho_s h_s \frac{\partial^2 w}{\partial t^2} = p^+ - p^-, \quad (14)$$

where D_s is the bending stiffness and ρ_s and h_s are, respectively, the mass density and the thickness of the panel, and ∇^4 is the biharmonic operator. Substituting Eq. (5) into Eq. (14) and using the orthogonality property of the structural modes, Eq. (14) is then transformed into a modal structural equation given by:

$$M_m (\omega_m^2 - \omega^2) q_m = \int_{A_f} p^+ \psi_m ds - \int_{A_f} p^- \psi_m ds, \quad (15)$$

where M_m is the structural generalized mass and ω_m the angular structural resonant frequency. In the real physical system, which is a combination of a series of subcavities, such a panel does not really exist. The purpose of the above formulation, using a panel, is just an artificial means to simulate the continuity between adjacent subcavities. Therefore, one can imagine a massless and stiffness-free membrane existing to separate the subcavities. As a special case of Eq. (15) by neglecting the generalized mass, the equation is reduced to

$$\int_{A_f} p^+ \psi_m ds - \int_{A_f} p^- \psi_m ds = 0. \quad (16)$$

For each subsystem (subcavity+membrane), the above equation will be used, together with the modal acoustic equation (9), to handle a more general case composed of a number of subsystems.

D. Generalization of the formulation

In order to generalize the procedure, the four subcavity system shown in Fig. 2(a) and (b) was first investigated. From the topological description illustrated in Fig. 2(b), Eq. (9) is applied to each subcavity k ($k = 1, 2, 3, 4$) and Eq. (16) for each membrane. This procedure results in the following

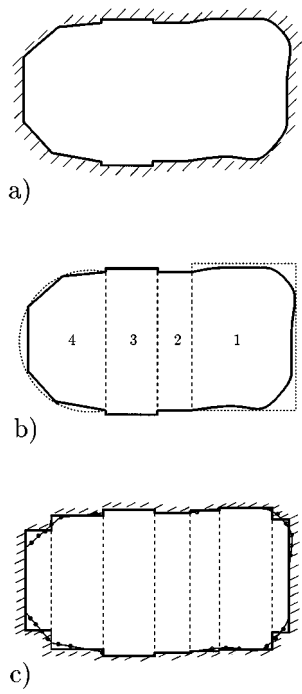


FIG. 2. Comparison between discretization procedures. (a) Real cavity; (b) discretization procedure used in the present approach; real cavity: —; bounding cavity:; panel: -----; (c) discretization procedure used in the acoustoelastic approach; real cavity: •••••; subcavities: —; panel: -----.

equations, expressed in matrix form, describing the coupling between the vectors P and U comprised respectively of the acoustic and structural modal amplitudes:

$$\begin{aligned} [A_k]\{P^k\} - [B_{k,k}]\{U^k\} &= 0, \quad k=1, \\ [A_k]\{P^k\} + [B_{k,k-1}]\{U^{k-1}\} - [B_{k,k}]\{U^k\} &= 0, \quad k=2,3, \\ [A_k]\{P^k\} + [B_{k,k-1}]\{U^{k-1}\} &= 0, \quad k=4, \end{aligned} \quad (17)$$

where the components of matrices $[A]$ and $[B]$ are

$$\begin{aligned} [A_k]_{ij} &= (\omega_{ik}^2 - \omega^2) \delta_{ij} + \frac{c^2}{\bigwedge_j V} \int_s \phi_j^k \frac{\partial \phi_i^k}{\partial \mathbf{n}_d^k}, \\ [B_{k,k}]_{ij} &= \omega^2 \frac{A_f^k}{V^k} [L_{k,k}]_{ij}, \\ [B_{k,k-1}]_{ij} &= \omega^2 \frac{A_f^{k-1}}{V^k} [L_{k,k-1}]_{ij}, \end{aligned}$$

where δ_{ij} is the Kronecker parameter and $[L]$ is the coupling matrix. Note that the first and the third equations hold for the two subcavities at the two ends with one separating membrane. The second equation is for the two intermediate subcavities which have two separating membranes at each end. On the other hand, the application of the modal structural equation Eq. (16) for each membrane leads to a set of equations summarized in the following matrix form:

$$[C_{k,k}]\{P^k\} - [C_{k+1,k}]\{P^{k+1}\} = 0, \quad k=1,2,3, \quad (18)$$

where the components of the matrix $[C]$ can be calculated from:

$$\begin{aligned} [C_{k,k}]_{ij} &= \frac{1}{\bigwedge_j^k} [L_{k,k}]_{ji}, \\ [C_{k+1,k}]_{ij} &= \frac{1}{\bigwedge_j^{k+1}} [L_{k+1,k}]_{ji}. \end{aligned}$$

Equation (18) describing the pressure continuity at the interface between adjacent portions can also be formulated in terms of the unknown vectors U containing the structural modal amplitudes. In fact, substituting Eqs. (17) into Eq. (18) gives

$$\begin{aligned} \sum_{j=0}^1 [-C_{k+j,k} A_{k+j}^{-1} B_{k+j,k}] \{U^k\} + [C_{k+1,k} A_{k+1}^{-1} B_{k+1,k+1}] \\ \times \{U^{k+1}\} &= 0, \quad k=1, \\ [C_{k,k} A_k^{-1} B_{k,k-1}] \{U^{k-1}\} + \sum_{j=0}^1 [-C_{k+j,k} A_{k+j}^{-1} B_{k+j,k}] \\ \times \{U^k\} + [C_{k+1,k} A_{k+1}^{-1} B_{k+1,k+1}] \{U^{k+1}\} &= 0, \quad k=2, \\ [C_{k,k} A_k^{-1} B_{k,k-1}] \{U^{k-1}\} + \sum_{j=0}^1 [-C_{k+j,k} A_{k+j}^{-1} B_{k+j,k}] \\ \times \{U^k\} &= 0, \quad k=3. \end{aligned} \quad (19)$$

In the more general case in which N subcavities are involved, this discretization procedure may be generalized. In fact, the governing equations for $k=1$ and $k=4$ can be directly applied to the first and the last subcavity of an arbitrary system, while the equations for $k=2,3$ hold for all intermediate subcavities. Furthermore, Eq. (18) applies to all membranes. The complete procedure yields the following general equation.

$$[Q]\{U\} = 0; \quad \{U\} = \{U^1, U^2, \dots, U^{N-1}\}^T. \quad (20)$$

Submatrices constituting the above system can be calculated as follows:

$$\begin{aligned} [Q_{i,j}] &= -[C_{i,i} A_i^{-1} B_{i,i}] - [C_{i+1,i} A_{i+1}^{-1} B_{i+1,i}], \quad i=j, \\ [Q_{i,j}] &= [C_{ji} A_j^{-1} B_{j,j}], \quad j=i+1, \\ [Q_{i,j}] &= [C_{ii} A_i^{-1} B_{i,j}], \quad i=j+1, \\ [Q_{i,j}] &= 0, \quad |i-j| \geq 2. \end{aligned} \quad (21)$$

It can be seen from these equations that the global $[Q]$ matrix is a banded one. This property is very helpful for reducing the final size of the matrix to be treated. In fact, using a partitioning technique of linear algebra for the banded matrix, any linear system can be reduced to a simple form involving only one set of modal coordinates. For example, if one chooses the coordinate set related to the first membrane $\{U^1\}$, successive substitutions yields

$$\begin{aligned} \{U^1\} &= [H_1]\{U^1\}, \quad \{U^2\} = [H_2]\{U^1\}, \\ \{U^k\} &= [H_k]\{U^1\}; \quad 3 \leq k \leq N-1, \end{aligned} \quad (22)$$

where

$$\begin{aligned} [H_1] &= [I], \quad [H_2] = -[Q_{1,2}^{-1}Q_{1,1}], \\ [H_k] &= -[Q_{k-1,k}^{-1}][Q_{k-1,k-2} + Q_{k-1,k-1}H_{k-1}], \\ 3 \leq k &\leq N-1. \end{aligned} \quad (23)$$

This whole procedure transforms the system expressed by Eq. (21) into a much more compact form using only the coordinates related to the first membrane:

$$\begin{aligned} \left[C_{k,k}A_k^{-1}B_{k,k-1}H_{k-1} - \sum_{j=0}^1 C_{k+j,k}A_{k+j}^{-1}B_{k+j,k}H_k \right] \{U^1\} \\ = 0, \quad k = N-1. \end{aligned} \quad (24)$$

Equation (24) should be solved to calculate the resonant frequencies. Note that Eq. (24) is a nonstandard eigenvalue problem so that numerical methods such as the bisection method should be used by making a distinction between poles and true solutions. The modal structural amplitude vector U^1 can then be calculated. A back substitution into Eq. (22) then gives all the other vectors U_k . Modal sound-pressure vectors P for each subcavity can then be computed from Eq. (17). For each resonant acoustic frequency ω_a , the modal shape of the cavity can also be evaluated using Eq. (4).

At this stage it is of interest to highlight the difference between the present formulation and the acoustoelastic method (AEM) concerning the discretization procedure. One particular aspect of the AEM, as illustrated in Fig. 2(c), is that the discretization introduces an approximation in the geometry of the irregular cavity, since the cavity, in practice, is replaced by a series of regular-shaped subcavities whose boundaries deviate more or less from the real cavity boundary. In the proposed integro-modal method (IMM), no such geometry approximation is necessary. Although the natural modes of the bounding cavity are used, it is the subcavities with real boundaries that are treated as shown in Fig. 2(b). From this point of view, the IMM is expected to be more accurate than the AEM while using fewer subcavities. On the other hand, the partitioning technique on the band matrix adopted here considerably reduces the size of the final matrix. Consequently, storage requirements are less demanding, permitting a substantial reduction in computation time.

II. NUMERICAL RESULTS

A. Remarks on numerical solution

The solution of Eq. (24) depends on the modal properties of the uncoupled component of each subsystem and the discretization of the whole enclosure. The truncation to a finite order of the decomposition series for both acoustic pressure and the membrane vibration and the number of subcavities used should be the two main factors affecting the accuracy of the method. Therefore, a convergence analysis of the method concerning these parameters is required. In order to ensure good convergence, three indices should be determined: the number of subcavities and the number of terms used in each of the acoustic and membrane series in Eqs. (4) and (5). For a given cavity, one can start with a small num-

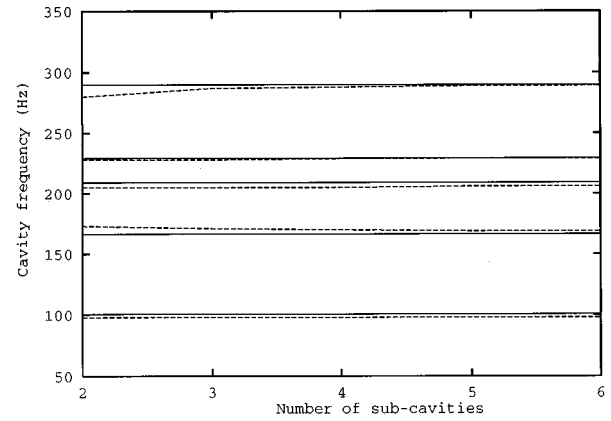


FIG. 3. Convergence study as a function of subcavity number for a semi-circular enclosure with unit radius. Exact: —; present approach: ---- with $n_a = 10$; $m_s = 10$.

ber of subcavities to get a rough idea about the frequency range of the modes of interest. The number of decomposition terms for each series can then be increased until no significant change is observed. Another cycle follows by increasing the number of subcavities and repeating the procedure for the decomposition series. This procedure should give the appropriate number of indices required to ensure a convergent stable solution. For a two-dimensional problem, two indices corresponding to the two orthogonal directions are involved for each acoustic mode. The maximum values of these indices are denoted by n_x and n_y . In all calculations reported hereafter, these two indices are always set equal; $n_x = n_y = n_a$. As far as the membrane is concerned, only one index is necessary, with m_s standing for the maximum number of terms used in the decomposition series.

In order to demonstrate the convergence properties of the method, a semi-circular cavity with a unit radius was investigated. In the calculations, rectangular bounding enclosures were used for each subcavity. The fluid speed inside the cavity was taken to be 343 m/s. The calculated natural frequencies are compared with the analytical solution given by:

$$\omega_{psq} = c \sqrt{\gamma_{ps}^2 + (q\pi/l)^2}, \quad (25)$$

where psq are the modal indices of the semi-cylindrical cavity of length l and radius a , γ_{ps} is the value of the s th root of the Bessel function of the first kind and order p : $J'_p(\gamma_{ps}a) = 0$. Since a two-dimensional problem is considered here, one has $q=0$.

The first example concerns convergence with respect to the number of subcavities, keeping the number of decomposition terms constant at an appropriate value (i.e., $n_a = 10$, meaning that 100 modes are used for each subcavity and 10 terms for each membrane). The convergence curves of the first five acoustic modes using the present method are compared with the exact solution in Fig. 3. It can be seen from Fig. 3 that the convergence rate depends on particular mode and that the computed solutions agree closely with the exact values. It is also worth noting that the solution seems to be

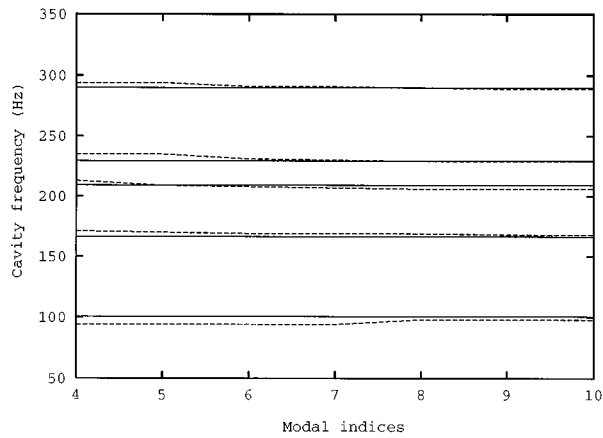


FIG. 4. Convergence study as a function of the modal indices of each subsystem. Exact: —; present approach: ----. The semi-circular enclosure is discretized by $N=5$ subcavities.

quite stable with a small number of subcavities, and relatively insensitive to any further increase in the number of subcavities.

Using five subcavities ($N=5$), the same investigation was carried out varying the number of terms in the different series. For the sake of convenience, the same values were taken for all the indices involved ($n_a=m_s$) and were increased accordingly. Comparisons with the analytical solution are presented in Fig. 4. Again, the calculated frequencies seem to converge quickly and closely to the exact solution with the increase in the number of decomposition terms.

A few remarks are necessary regarding the numerical technique for the calculation of the natural frequencies. As mentioned before, the bisection method is used in the present work by making a distinction between poles and true solutions to find the roots of the characteristic equation of the nonstandard eigenvalue problem. This procedure is illustrated in Fig. 5 based on the cavity treated previously. In this

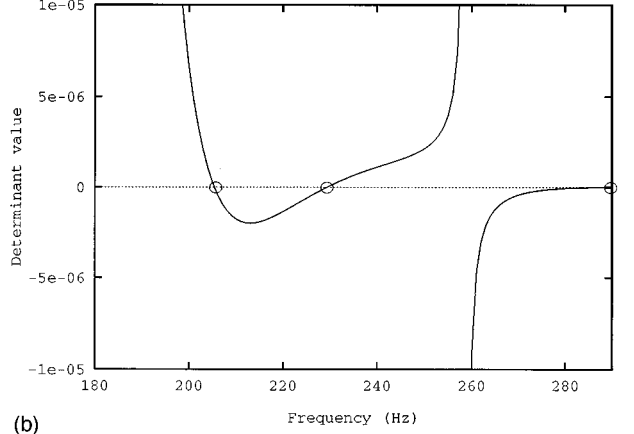
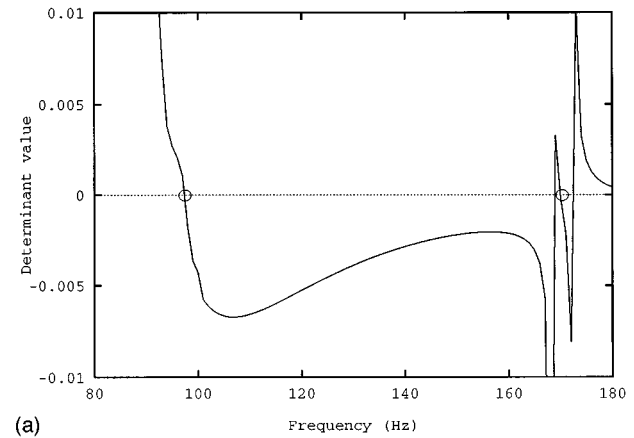


FIG. 5. Graphical representation of computed solutions of a semi-circular enclosure with $N=5$; $n_a=10$; $m_s=10$. (a) Frequency range 80–180 Hz; (b) frequency range 180–290 Hz.

figure, the determinant of Eq. (24) is plotted in Fig. 5(a) and 5(b) over a frequency band ranging from 80 to 290 Hz, comprising the five first modes. As can be seen, the presence of poles changes locally the property of the characteristic equa-

TABLE I. Resonant frequencies of a semi-circular enclosure with unit radius: A comparison between the present approach ($N=5$; $n_a=m_s=10$), the acoustoelastic method ($N=15$; $n_a=m_s=10$) and the exact solution.

| Mode order | Exact solution f_n (Hz) | Present approach | | Acoustoelastic method | |
|------------|------------------------------|------------------|-----------|-----------------------|-----------|
| | | f_n (Hz) | Error (%) | f_n (Hz) | Error (%) |
| 1 | 100.4459 | 98.0 | -2.43 | 100.0 | -0.44 |
| 2 | 166.4999 | 168.0 | 0.90 | 165.0 | -0.90 |
| 3 | 209.0803 | 206.0 | -1.47 | 210.0 | 0.44 |
| 4 | 229.2786 | 229.0 | -0.12 | 229.0 | -0.12 |
| 5 | 289.8737 | 289.0 | -0.30 | 287.0 | -0.99 |
| 6 | 290.9655 | 295.0 | 1.38 | 293.0 | 0.70 |
| 7 | 349.9228 | 354.0 | 1.16 | 346.0 | -1.12 |
| 8 | 365.7540 | 369.0 | 0.88 | 367.0 | 0.34 |
| 9 | 382.6769 | 383.0 | 0.08 | 386.0 | 0.86 |
| 10 | 409.4261 | 412.0 | 0.63 | 404.0 | -1.32 |
| 11 | 437.2671 | 437.0 | -0.06 | 440.0 | 0.62 |
| 12 | 465.6539 | 463.0 | -0.57 | 457.0 | -1.85 |
| 13 | 467.8375 | 475.0 | 1.53 | 468.0 | 0.03 |
| 14 | 506.5966 | 510.0 | 0.67 | 508.0 | 0.27 |
| 15 | 526.2490 | 532.0 | 1.09 | 519.0 | -1.37 |
| 16 | 543.7178 | 544.0 | 0.05 | 546.0 | 0.42 |
| 17 | 555.1818 | 554.0 | -0.21 | 559.0 | 0.68 |
| 18 | 573.7424 | 567.0 | -1.17 | 565.0 | -1.52 |
| 19 | 584.6605 | 583.0 | -0.28 | 576.0 | -1.48 |

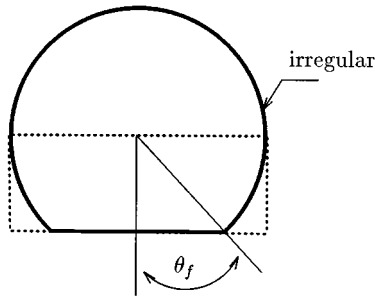


FIG. 6. A two-dimensional simplified aircraft cabin: —; bounding cavity:

tion. These poles correspond to subcavity frequencies which make the matrix $[A_k]$ singular so that they are not the true resonant frequencies of the whole cavity. In our case, for example, the first solution is a true root, while the second one is a pole. For the whole frequency range under consideration, all five roots are marked by a circle in the various plots [Fig. 5(a) and (b)].

B. Validation and comparison with other methods

Validation tests were also performed using the regular cavity cited above and another geometry simulating a simplified two-dimensional aircraft cabin, including comparisons with other available methods. All the simulations used $n_a = m_s = 10$ with a variable number of subcavities.

1. Semi-circular cavity

Again, the semi-circular cavity used before was investigated. The computed results of cavity frequencies for a frequency range of 90–600 Hz are reported in Table I, together with a comparison between the present method (IMM), the acoustoelastic method (AEM), and the analytical solution. It can be observed that both methods agree well with the analytical solutions. The mean percentage error (calculated with absolute values) over the whole frequency range is about 0.8% for both methods. However, only 5 subcavities were used in the present method while 15 subcavities were used in AEM. Also, the present approach needs 30% of the discretization size of the AEM. From this point of view, the proposed approach seems to be more efficient than the AEM to achieve comparable accuracy.

2. Two-dimensional simplified aircraft cabin

Another computational example was performed on a simplified aircraft cabin with an irregular shape. This configuration is of great interest in recent research on aircraft cabin noise.^{9–11} The addition of the floor to the cylindrical model leads to a more realistic configuration compared with the single cylindrical model reported frequently in the literature, but results in a cavity cannot be treated by analytical methods. The cavity geometry is illustrated in Fig. 6. The proposed method is applied to this configuration which is discretized into two subcavities: an upper semi-circular cavity coupled to a lower irregular cavity with a rectangular envelope. It is equally possible to use a semi-circular enve-

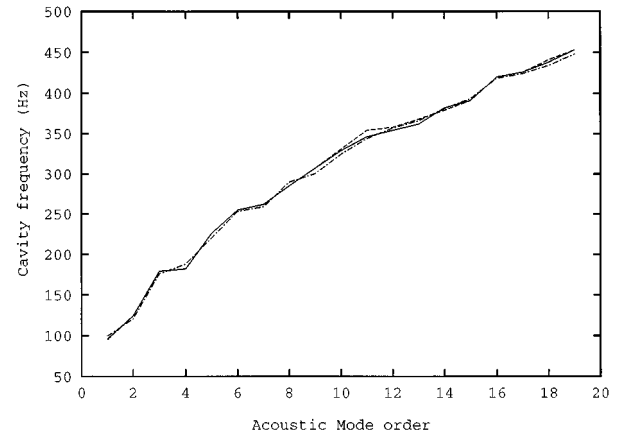


FIG. 7. Acoustic frequencies of a two-dimensional simplified aircraft cabin with unit radius and a floor location of $\theta_f = 56.6^\circ$. Frequencies calculated using the present approach with $N=2$: - · - · - ·; the acoustoelastic method with $N=5$: ----; the finite difference method in Ref. 12: —.

lope for the lower subcavity. The acoustoelastic coupling between the semi-circular cavity and the membrane is evaluated by a numerical integration over the vibrating surface. The modal coupling coefficient is given by

$$L_{psm} = \frac{1}{A_f} \int_0^a J_p(\gamma_{ps}r) \left[\sin\left(\frac{m\pi(r+a)}{2a}\right) + (-)^p \sin\left(\frac{m\pi(a-r)}{2a}\right) \right] dr, \quad (26)$$

where (p,s) are the acoustic modal indices and a is the radius of the semi-circle. With an angle of $\theta_f = 56.6^\circ$ defining the position of the floor, different results regarding the natural frequencies are compared in Fig. 7. The results used were calculated respectively by IMM, AEM, and also taken from Ref. 12 based on the finite difference approach. The IMM used two subcavities while the AEM used five. The agreement between the three methods is reasonably good. Once again, comparable accuracy was obtained using IMM, with less subcavities than the AEM.

One of the appealing features of the IMM is its flexibility in choosing different strategies to handle different configurations. As an example, consider the cabin configuration

TABLE II. Resonant frequencies of a simplified cabin with a floor location of $\theta_f = 49^\circ$ and a unit radius: A comparison between the present approach ($N=1$; $n_a=10$) and Ref. 13.

| Mode order | Reference 13 f_n (Hz) | Present approach | |
|------------|----------------------------|------------------|-----------|
| | | f_n (Hz) | Error (%) |
| 1 | 96.416 | 95.0 | −1.47 |
| 2 | 113.002 | 117.0 | 3.53 |
| 3 | 166.997 | 173.0 | 3.60 |
| 4 | 178.748 | 179.0 | 0.14 |
| 5 | 212.718 | 217.0 | 2.01 |
| 6 | 238.019 | 244.0 | 2.51 |
| 7 | 243.915 | 252.0 | 3.31 |
| 8 | 280.915 | 283.0 | 0.74 |
| 9 | 283.663 | 288.0 | 1.53 |
| 10 | 305.542 | 295.0 | −3.45 |

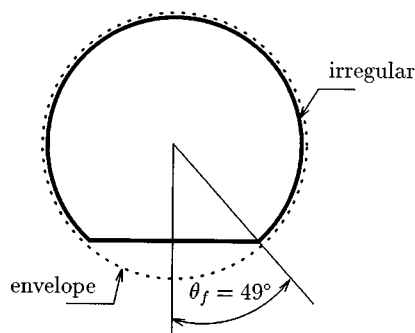


FIG. 8. Discretization of a two-dimensional simplified aircraft cabin of unit radius using one subcavity. The floor location is at $\theta_f = 49^\circ$.

with $\theta_f = 49^\circ$ and unit radius, as in Fig. 8. The fact that the floor position is lower than the previous one makes the cavity closer to a completely circular one. In this case, the enclosure can be treated as one single cavity with a complete circular envelope. In this case, only Eq. (9) is needed with the left hand side term equal to zero, since no membrane is present in the system. A comparison with Ref. 13 is given in Table II. It can be observed that, although the agreement is acceptable, the error is greater than that obtained previously. This can certainly be attributed to the deviation in the boundary shape, since even with the low floor position it is still quite different from the envelope. Accuracy can certainly be enhanced by using more subcavities at the price of increasing the calculation effort.

III. EXPERIMENTAL VALIDATION

Experimental tests were performed to assess the proposed method on a cavity simulating an aircraft cabin. The experimental setup and instrumentation used are illustrated in Fig. 9. The cavity was formed by a steel cylinder with a floor. The interior space of concern was the volume above the floor. The ends of the cavity were closed with thick steel end caps. The test cylinder had an internal diameter of 0.504 m and was 1.1684 m long. The floor was made of the same material as the cylinder and was located at an angle θ_f of 49° . Two $\frac{1}{4}$ -in. microphones were placed inside the cavity supported by a thin tube along the cylinder centerline. The tube could be rotated and moved along the centerline to pick

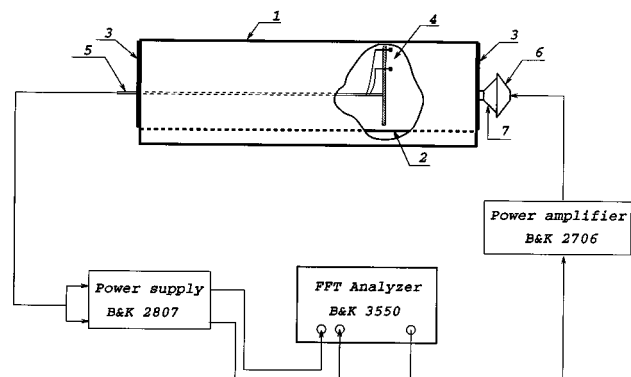


FIG. 9. Experimental setup: (1): cylinder; (2): floor; (3): end caps; (4): $\frac{1}{4}$ -in. microphones; (5): tube; (6): loudspeaker; (7): connector.

TABLE III. Resonant frequencies of a simplified cabin with a floor location of $\theta_f = 49^\circ$ and a 0.254 m radius. A comparison between the present approach ($N=2$; $n_a=m_s=10$) and experimental results.

| Mode order | Experimental results f_n (Hz) | Present approach | |
|------------|------------------------------------|------------------|-----------|
| | | f_n (Hz) | Error (%) |
| 1 | 380.0 | 375.0 | -1.31 |
| 2 | 460.0 | 463.0 | 0.65 |
| 3 | 678.0 | 684.0 | 0.88 |
| 4 | 720.0 | 710.0 | -1.39 |
| 5 | 848.0 | 856.0 | 0.94 |
| 6 | 968.0 | 963.0 | -0.51 |

out any desired measurement point. Acoustic excitation was produced by a loudspeaker with a cone connector fixed to the right end cap through a hole. The characteristics of the cone connector was chosen in such a way that neither its acoustic nor its structural resonances lie in the frequency range of interest. The measured sound pressure was treated by a multi-channel BK 3550 FFT analyzer. Care was taken to avoid any strong coupling between the cavity and the boundary formed by structure. It was observed that stiffening the structure did not noticeably change the measured resonant frequencies. Although more than a dozen resonant modes were clearly identified, only the two-dimensional modes, corresponding to modes in the cross section are compared with our calculations in Table III. The values of these two-dimensional modes can be used to derive the other three-dimensional modes involved, since simple formulas are available for the longitudinal direction. It can be seen that the theoretical results and the experimental measurements agree well with a maximal error of 1.3%.

IV. CONCLUSION

A new approach has been proposed for the computation of acoustic modes of irregular-shaped cavities. The method approximates the solution via an integro-modal formulation using multi-connected subcavities. The formulation is general and flexible enough to handle different cavity configurations. From this point of view, the Green's function method can be considered as a special case of the present technique using one single bounding cavity. With respect to the acoustoelastic method, the proposed formulation permits the use of irregular-shaped subcavities, thus making the approach more powerful. Numerical results on two-dimensional cavities have been presented to demonstrate the efficiency and the accuracy of the approach. Preliminary comparisons with the acoustoelastic method shows that the proposed technique gives comparable results using fewer subcavities.

Future work is required to extend this approach to the prediction of interior noise inside irregular-shaped cavity coupled with vibrating structures.

¹M. Petyt, J. Lea, and G. H. Koopman, "A finite element method for determining the acoustic modes of irregular shaped cavities," *J. Sound Vib.* **45**, 497-502 (1976).

²P. D. Joppa and I. M. Fyfe, "A finite element analysis of the impedance properties of irregular shaped cavities with absorptive boundaries," *J. Sound Vib.* **56**, 61-69 (1978).

- ³T. L. Richards and S. K. Jha, "A simplified finite element method for studying acoustic characteristics inside a car cavity," *J. Sound Vib.* **63**, 61–72 (1979).
- ⁴J. F. Unruh, "A finite element subvolume technique for structural-borne interior noise prediction," *J. Aircraft* **17**, 434–441 (1980).
- ⁵E. H. Dowell, G. F. Gorman III, and D. A. Smith, "Acoustoelasticity: General theory, acoustic natural modes and forced response to sinusoidal excitation, including comparison with experiment," *J. Sound Vib.* **52**, 519–542 (1977).
- ⁶C.-F. Chao, E. H. Dowell, and D. B. Bliss, "Modal analysis of interior noise fields," in *Proceedings of the Vibrations Conference of the Design Engineering Technical Conferences* (ASME, New York, 1981), pp. 30–56.
- ⁷P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Vol. II (McGraw-Hill, New York, 1953).
- ⁸G. P. Succi, "The interior acoustic field of an automobile cabin," *J. Acoust. Soc. Am.* **81**, 1688–1694 (1987).
- ⁹C. R. Fuller, "Structural influence of cabin floor on sound transmission into propeller aircraft-analytical investigation," AIAA Paper 86-1940 (1986).
- ¹⁰L. D. Pope, "On the prediction of propeller tone sound levels and gradients in an airplane cabin," *J. Acoust. Soc. Am.* **88**, 2755–2765 (1990).
- ¹¹J. Missaoui, L. Cheng, and M. J. Richard, "Free and forced vibration of a cylindrical shell with a floor partition," *J. Sound Vib.* **190**, 21–40 (1996).
- ¹²L. D. Pope, E. G. Wilby, and J. F. Wilby, "Propeller aircraft interior noise model," NASA CR 3813 (1984).
- ¹³L. D. Pope, E. G. Wilby, and J. F. Wilby, "Analytical predictor of the interior noise for cylindrical models of aircraft fuselage for prescribed exterior noise fields. Phase II: Models for sidewall trim, stiffened structures, and cabin acoustics with floor partition," NASA CR 165869 (1982).

Effects of an elastic membrane on tube waves in permeable formations

Hsui-Lin Liu and David Linton Johnson

Schlumberger-Doll Research, Old Quarry Road, Ridgefield, Connecticut 06877-4108

(Received 9 June 1996; revised 7 October 1996; accepted 20 November 1996)

The properties of tube wave (Stoneley wave) propagation in a fluid-filled borehole penetrating a permeable rock are modified by the presence of a mudcake which may form on the borehole wall. A theoretical description of these effects characterizes the bulk properties of the porous rock using the Biot theory and the mudcake as an elastic impermeable layer. A finite membrane stiffness simulates the mudcake partial sealing mechanism. Within the context of this model, the mudcake can reduce, but not eliminate, the permeability effects on the tube wave slowness and attenuation. Numerical examples relevant to a typical sandstone of porosity 20% and permeability 200 mD, show that (a) the membrane stiffness acts to reduce the permeability effect on slowness and on attenuation and (b) a finite thickness of mudcake shifts the phase slowness by a (nearly) frequency-independent amount, with little effect upon the attenuation. An expression for the low-frequency limit clarifies the important combinations of mudcake parameters. © 1997 Acoustical Society of America. [S0001-4966(97)03705-3]

PACS numbers: 43.20.Mv, 43.20.Ks, 43.20.Gp, 43.40.Dx [JEG]

INTRODUCTION

In the last 20 or so years there has been a continuing effort in attempting to quantify the correlations found between the properties of a borehole tube wave (often referred to as a Stoneley wave) and the fluid flow permeability of the rock formation, using a sonic tool (Rosenbaum, 1974; Williams *et al.*, 1984; Cheng *et al.*, 1987; Schmitt *et al.*, 1988). In laboratory experiments, Winkler *et al.* (1989) used scale modeled boreholes to demonstrate that the attenuation and the dispersion of the borehole tube wave in permeable rocks free of mudcake can accurately be calculated by the Biot theory (1956a, b) applied to this cylindrical geometry. There are no adjustable parameters in this theory. The potential application of this technology, however, is to oil field acoustic logging in situations in which a mudcake layer may have developed on the borehole wall. This mudcake has two general effects on tube wave characteristics: Its elastic characteristics are different from that of the borehole fluid, thus changing the overall slowness of the mode, and it acts to partially seal the oscillatory flow of fluid between the borehole and the formation, thus reducing the effect of formation permeability.

In this article we focus our attention upon these two effects. Our model of the mudcake is that it is an impermeable elastic layer. We consider the effects of a membrane stiffness which allows for a partial flexing of the mudcake into the pores of the formation; this stiffness enters via the relevant boundary condition (see Sec. I) between the mudcake and the porous solid. As before, (Winkler *et al.*, 1989) we treat the formation properties within the context of the Biot theory. In terms of the elastic parameters and the thickness of the layer, and the stiffness of the membrane, we are able to calculate the effects of the mudcake on tube wave slowness and attenuation. Our model is similar to, but different from, that of Tang and Martin (1994), who did not specifically include the effects of a finite thickness of mudcake,

and that of Schmitt (1988) who considered the effects of a permeable layer having a different permeability than the formation. Schmitt did not consider the effects of an elastic layer having zero permeability as we have done. Had he done so, his model would be the same as ours except that the membrane stiffness (Sec. I) would be infinite in value, which we believe to be unphysical.

The outline of this article is as follows. In the next section we develop the theoretical treatment of tube wave characteristics in the presence of a mudcake layer lining a permeable formation. We present numerical results for the complex-valued wave number $k_z(\omega)$ using a few relevant parameter sets in Sec. II. These characteristics are determined by solving for the complex valued roots of a specific 8×8 determinant. In general, this must be done numerically. When certain low-frequency approximations are appropriate, it is possible to derive a closed-form expression for $k_z(\omega)$, and this is done in Sec. III. The significance of this result is that it makes clear just what combination of mudcake parameters is important. Our concluding remarks appear in Sec. IV.

I. THEORY

A model illustrating the borehole configuration is shown in Fig. 1. An elastic and flexible mudcake layer, of inner radius a and outer radius b , is situated between the borehole fluid and the formation solid. The mudcake flexibility is introduced by adding a membrane stiffness on the borehole wall to allow the membranelike mudcake to flex in and out of the pore space. This mechanism reduces, but does not eliminate, the effects of formation permeability on the tube wave.

To characterize the tube wave properties, we search for the axially symmetric normal modes which vary as $e^{i(k_z z - \omega t)}$ in a fluid-filled cylindrical borehole surrounded by porous rock. Here, z indicates the vertical axis and k_z is the

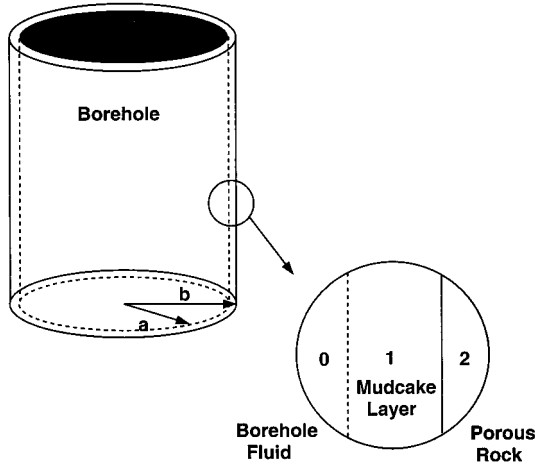


FIG. 1. Specified boundaries in a borehole.

axial wave number, a complex-valued function of frequency. The rock bulk properties are characterized using the Biot theory (1956a,b). We use the formulation discussed by Johnson *et al.* (1994a,b) where it is shown that all of the parameters entering into the Biot theory can be measured by independent means. The properties of the fast, slow, and shear waves can then be calculated and compared directly against the experimentally measured counterparts; the Biot theory is shown to yield an excellent quantitative description of the speeds and attenuations of these modes. We use the dynamic permeability derived by Johnson *et al.* (1987) to specify the frequency-dependent nature of the rock permeability. The boundary conditions derived by Deresiewicz and Skalak (1963) and employed by Rosenbaum (1974) are used here. They are more general than the specific elastic-porous boundary conditions reviewed by Schmitt *et al.* (1988). We briefly summarize these boundary conditions below.

Let U denote the fluid displacement, u the solid displacement, P the fluid pressure, and τ_{ij} the stress component on the solid phase; the continuity conditions at the boundary between the borehole fluid and the elastic mudcake layer are

$$U_{r0}(a^-) = u_{r1}(a^+), \quad (1)$$

$$P_0(a^-) = -\tau_{rr1}(a^+), \quad (2)$$

and

$$0 = \tau_{rz1}(a^+). \quad (3)$$

The cylindrical coordinates r and z denote the radial and the vertical axes. The subscripts 0 and 1 indicate the borehole fluid and the mudcake layer, respectively.

At the boundary of the elastic mudcake layer and the porous rock, we allow the mud layer to flex like a membrane across the pore opening. This leads to the following constraints: continuity of fluid volume:

$$u_{r1}(b^-) = (1 - \phi)u_{r2}(b^+) + \phi U_{r2}(b^+), \quad (4)$$

continuity of axial displacement (solid phase):

$$u_{z1}(b^-) = u_{z2}(b^+), \quad (5)$$

continuity of total stress:

$$\tau_{rr1}(b^-) = \tau_{rr2}(b^+) - \phi P_2(b^+), \quad (6)$$

$$\tau_{rz1}(b^-) = \tau_{rz2}(b^+), \quad (7)$$

where τ_{ij2} refers to the stress on the solid component of the porous medium, and membrane flexural response,

$$-\tau_{rr1}(b^-) = P_2(b^+) + W_{mc}\phi[U_{r2}(b^+) - u_{r2}(b^+)]. \quad (8)$$

ϕ is the rock porosity, ω is the angular frequency, and W_{mc} is the mudcake membrane stiffness. The subscript 2 indicates the porous rock.

Boundary conditions Eqs. (1)–(8) represent our model of the borehole acoustics. All are fairly obvious except Eq. (8), which can be understood as follows: When the thickness of the mudcake reduces to zero, condition Eq. (8) becomes

$$P_0 - P_2 = W_{mc}\phi(U_{r2} - u_{r2}), \quad (9)$$

which is the same as that described by Rosenbaum (1974); $W_{mc} = -i\omega\kappa(\omega)$ in his notation. In the present article, we model the mudcake as a layer which can flex in and out of the pores of the formation as the local pressure oscillates up and down. W_{mc} describes the relative displacement between the pore fluid and the solid frame at the mudcake-formation boundary in terms of the difference in stress across that boundary. Thus, in our numerical examples we consider cases for which W_{mc} , not κ , is real-valued. Tang (1994) has developed a theory of W_{mc} in the specific case that the porous medium consists of cylindrical holes of radius R_0 ; he finds $W_{mc} = 3\pi\mu/4R_0\phi(1 - \nu)$ where μ and ν are the shear modulus and Poisson's ratio, respectively, of the mudcake material and ϕ is the porosity of the formation. Similarly, Nagy and Blaho (1994) have considered the membrane effect between two different fluids caused by surface tension, for which $W_{mc} \propto 1/R_0^2$. Here, we take W_{mc} as a parameter of the problem. We note that the specific cases of an open and an epoxy-sealed interface were shown to be accurately described by $W_{mc} = 0$ and $W_{mc} \rightarrow \infty$, respectively, in reflectivity experiments reported in Johnson *et al.* (1994b).

Our model is somewhat analogous to that of Schmitt (1988) who considered the effects of a *permeable* layer, which may be either a mudcake or a damaged zone. In order to make contact with the model of the present paper, the permeability of that layer would have to be taken as zero. In this limit, however, the Schmitt model disallows the possibility of relative motion between solid and fluid at the formation wall: $[U_{r2}(b^+) - u_{r2}(b^+)] \equiv 0$. This, then, becomes a special case of our model but with $W_{mc} \rightarrow \infty$ [cf. Eq. (8)]. We consider the effects of a nonzero layer thickness and a finite membrane stiffness to be important, and the permeability of that layer to be negligible, for reasons given in the next section.

In order to complete the description of the borehole acoustics, we need six displacement potentials to describe the wave fields. By omitting the common term of $e^{ik_z z} e^{-i\omega t}$, these potentials can be specified by the eight unknown amplitude functions following Chang *et al.* (1988) (see also Kurkjian and Chang, 1986 for more details).

The displacement potential of the borehole fluid is,

$$\Phi_0 = H_0^{(1)}(k_{r0}r) + A_0(k_z, \omega)J_0(k_{r0}r), \quad (10)$$

in which the first term represents the monopole source at the origin and the second represents the change induced by the presence of the mudcake and formation. In the elastic mudcake, we have the compressional displacement potential,

$$\Phi_1 = B_1(k_z, \omega) H_0^{(1)}(k_{r1_c} r) + A_1(k_z, \omega) J_0(k_{r1_c} r); \quad (11)$$

and the shear displacement potential,

$$\Psi_1 = D_1(k_z, \omega) H_1^{(1)}(k_{r1_{sh}} r) + C_1(k_z, \omega) J_1(k_{r1_{sh}} r). \quad (12)$$

In the permeable medium, we have three displacement potentials:

Fast wave,

$$\Phi_{2F} = A_2(k_z, \omega) H_0^{(1)}(k_{r2_F} r); \quad (13)$$

slow wave,

$$\Phi_{2S} = B_2(k_z, \omega) H_0^{(1)}(k_{r2_S} r); \quad (14)$$

shear wave,

$$\Psi_2 = C_2(k_z, \omega) H_1^{(1)}(k_{r2_{sh}} r). \quad (15)$$

$H_0^{(1)}$, $H_1^{(1)}$, J_0 , J_1 are Hankel and Bessel functions. The subscripts 0, 1, and 2 of the amplitudes A_0 , A_1 , B_1 , C_1 , D_1 , A_2 , B_2 , C_2 denote, respectively, the borehole fluid, the elastic mudcake, and the surrounding permeable medium. F , S , and sh indicate the fast wave, the slow wave, and the shear wave of the permeable medium. k_r is the radial wave number, and relates to the material property by $k_r^2 = (\omega/v)^2 - k_z^2$, where $v(\omega)$ is the complex-valued phase velocity of the bulk wave (fast, slow, or shear) in question.

The displacements and stresses can be expressed in terms of the displacement potentials. In the fluid-filled borehole, the radial displacement is

$$U_{r0} = \frac{\partial \Phi_0}{\partial r}, \quad (16)$$

and the pressure in the borehole fluid is

$$P_0 = -\tau_{rr0} = \rho_{f0} \omega^2 \Phi_0. \quad (17)$$

For the elastic mudcake, we have the following expressions for the radial displacement,

$$u_{r1} = \frac{\partial \Phi_1}{\partial r} - \frac{\partial \Psi_1}{\partial z}; \quad (18)$$

the vertical displacement,

$$u_{z1} = \frac{\partial \Phi_1}{\partial z} + \frac{\partial \Psi_1}{\partial r} + \frac{\Psi_1}{r}; \quad (19)$$

the normal stress,

$$\tau_{rr1} = (\lambda + 2\mu) \frac{\partial u_{r1}}{\partial r} + \mu \left(\frac{u_{r1}}{r} + \frac{\partial u_{z1}}{\partial z} \right); \quad (20)$$

and the tangential stress,

$$\tau_{rz1} = \mu \left(\frac{\partial u_{z1}}{\partial r} + \frac{\partial u_{r1}}{\partial z} \right). \quad (21)$$

The Biot theory (Biot, 1956a,b) is used to describe the bulk properties of the permeable medium. We follow the derivation by Chang *et al.* (1988) to compute the fast wave, slow wave, and shear wave contributions to the displacements in the formation. Using cylindrical coordinates, the displacements relate to the displacement potentials by the following expressions:

Radial displacement of the solid,

$$u_{r2} = \frac{\partial \Phi_{2F}}{\partial r} + \frac{\partial \Phi_{2S}}{\partial r} - \frac{\partial \Psi_2}{\partial z}; \quad (22)$$

vertical displacement of the solid,

$$u_{z2} = \frac{\partial \Phi_{2F}}{\partial z} + \frac{\partial \Phi_{2S}}{\partial z} + \frac{\partial \Psi_2}{\partial r} + \frac{\Psi_2}{r}; \quad (23)$$

radial displacement of the pore fluid,

$$U_{r2} = G_+ \frac{\partial \Phi_{2F}}{\partial r} + G_- \frac{\partial \Phi_{2S}}{\partial r} - G_{sh} \frac{\partial \Psi_2}{\partial z}; \quad (24)$$

where

$$G_+ = -\frac{v_{c1}^2 \tilde{\rho}_{11} - P}{v_{c1}^2 \tilde{\rho}_{12} - Q}, \quad (25)$$

$$G_- = -\frac{v_{c2}^2 \tilde{\rho}_{11} - P}{v_{c2}^2 \tilde{\rho}_{12} - Q}, \quad (26)$$

and

$$G_{sh} = \frac{\tilde{\alpha} - 1}{\tilde{\alpha}}. \quad (27)$$

The stresses in the permeable medium can be expressed as:

Normal stress on the solid phase,

$$\tau_{rr2} = (P - 2N) d_{2s} + Q d_{2f} + 2N \frac{\partial u_{r2}}{\partial r}; \quad (28)$$

tangential stress on the solid phase,

$$\tau_{rz2} = N \left(\frac{\partial u_{r2}}{\partial z} + \frac{\partial u_{z2}}{\partial r} \right); \quad (29)$$

pressure of the pore fluid,

$$P_2 = \frac{-1}{\phi} (R d_{2f} + Q d_{2s}); \quad (30)$$

where

$$d_{2s} = \frac{\partial u_{r2}}{\partial r} + \frac{u_{r2}}{r} + \frac{\partial u_{z2}}{\partial z}, \quad (31)$$

and

$$d_{2f} = \frac{\partial U_{r2}}{\partial r} + \frac{U_{r2}}{r} + \frac{\partial U_{z2}}{\partial z}. \quad (32)$$

v_{sh} , v_{c1} , and v_{c2} denote the shear, the fast wave, and the slow wave velocities of the permeable medium. $\tilde{\alpha}$ is the complex tortuosity. A thorough description of its definition and nature can be found in Johnson *et al.* (1987), and it can be accurately approximated as,

$$\tilde{\alpha}(\omega) = \alpha_{\infty} + \frac{i\eta\phi}{\omega\kappa_0\rho_f} \left[1 - \frac{4i\alpha_{\infty}^2\kappa_0^2\rho_f\omega}{\eta\Lambda^2\phi^2} \right]^{1/2}, \quad (33)$$

where η is the pore fluid viscosity, κ_0 (not to be confused with the surface flow impedance κ discussed above) is the dc permeability, α_{∞} is the tortuosity at infinite frequency, and Λ is a measure of dynamically connected pore size. In the present article we presume the approximate validity of the relationship $\Lambda^2 = 8\alpha_{\infty}\kappa_0/\phi$; see Johnson *et al.* (1994a, b) and references therein. The complex permeability can easily be calculated from the complex tortuosity by $\tilde{\kappa}(\omega) = i\eta\phi/\tilde{\alpha}(\omega)\omega\rho_f$.

P , Q , R , and N are the general elastic coefficients. These coefficients relate to the material properties as follows: Shear wave velocity,

$$v_{sh}^2 = \frac{N}{[(1-\phi)\rho_s + (1-\tilde{\alpha}^{-1})\phi\rho_f]}; \quad (34)$$

fast wave and slow wave velocities,

$$v_{(c1,c2)}^2 = \frac{\Delta \pm [\Delta^2 - 4(\tilde{\rho}_{11}\tilde{\rho}_{22} - \tilde{\rho}_{12}^2)(PR - Q^2)]^{1/2}}{2(\tilde{\rho}_{11}\tilde{\rho}_{22} - \tilde{\rho}_{12}^2)}, \quad (35)$$

where

$$\Delta = P\tilde{\rho}_{22} + R\tilde{\rho}_{11} - 2Q\tilde{\rho}_{12}, \quad (36)$$

$$\tilde{\rho}_{11} + \tilde{\rho}_{12} = (1-\phi)\rho_s, \quad (37)$$

$$\tilde{\rho}_{22} + \tilde{\rho}_{12} = \phi\rho_f, \quad (38)$$

$$\tilde{\rho}_{12} = -(\tilde{\alpha} - 1)\phi\rho_f, \quad (39)$$

$$P = (1-\phi)Q/\phi + K_b R/(\phi K_f) + 4N/3, \quad (40)$$

$$Q = \frac{\phi K_s(1-\phi-K_b/K_s)}{1-\phi-K_b/K_s + \phi K_s/K_f}, \quad (41)$$

$$R = \frac{\phi^2 K_s}{1-\phi-K_b/K_s + \phi K_s/K_f}. \quad (42)$$

ρ_s , ρ_f are the solid and pore fluid densities ϕ is the porosity. K_s , K_f , K_b , and N denote the solid grain modulus, the pore fluid bulk modulus, the frame bulk modulus, and the frame shear modulus, respectively. These quantities are measurable for a porous medium and shall be input to the computation.

By applying the eight boundary conditions of Eqs. (1)–(8), we obtain eight governing equations which can be put in the form $[M]\mathbf{A}=\mathbf{B}$ where the vector \mathbf{A} consists of the eight unknown amplitudes: A_0 , A_1 , B_1 , C_1 , D_1 , A_2 , B_2 , C_2 . $[M]$ consists of 8×8 matrix elements of the material properties and \mathbf{B} has the components of the source terms. A normal mode is a nontrivial solution which can exist in the absence of an external source, i.e., when $\mathbf{B}=0$. At each frequency, the mode is defined by that k_z which ensures $\det[M]=0$. These roots are determined by a numerical search algorithm as before (Chang *et al.*, 1988).

II. NUMERICAL RESULTS

We present results based on calculations for a more or less typical sandstone formation. All parameter values are

TABLE I. Values of input parameters for the calculation of tube wave characteristics.

| Borehole parameters | ρ_{f0} (g cm ⁻³) K_{f0} (dyn cm ⁻²) | 1.0 2.25×10^{10} |
|----------------------|--|---|
| | ρ_{s1} (g cm ⁻³) | 1.1 |
| Mudcake parameters | λ (dyn cm ⁻²) μ (dyn cm ⁻²) $[b-a]$ (mm) W_{mc} (GPa/cm) | 2.59×10^{10} 1.13×10^9 [0–4] [0–10] |
| | b (cm) ϕ α_{∞} κ_0 (μm) ² Λ (μm) | 10.5 0.20 2.24 [0–1] $\sqrt{8\alpha_{\infty}\kappa_0/\phi}$ |
| Formation parameters | ρ_f (g cm ⁻³) K_f (dyn cm ⁻²) η (poise) ρ_s (g cm ⁻³) K_s (dyn cm ⁻²) K_b (dyn cm ⁻²) N (dyn cm ⁻²) | 1.0 2.25×10^{10} 0.010 2.67 3.5×10^{11} 1.61×10^{11} 5.59×10^{10} |

collected in Table I. We take the unit of permeability 1 darcy = $1(\mu\text{m})^2$, exactly. We have assumed that the dry speed for the compressional wave is $V_p = 3360$ m/s and for the shear $V_s = 1675$ m/s, from which we deduce values for K_b and N . As mentioned earlier, we have assumed the approximate validity of the relationship $\Lambda = \sqrt{8\alpha_{\infty}\kappa_0/\phi}$, which has been verified by means of numerical calculations on a wide variety of pore geometries and has been established experimentally on a number of samples (Johnson *et al.*, 1994a, b) and references therein]. No intrinsic attenuation is included for the solid frame, although it could easily be included by using complex-valued frame moduli. The borehole fluid is presumed to be water. We relate the phase slowness $S(\omega)$ and the specific attenuation $Q^{-1}(\omega)$ to the complex-valued wave number:

$$k_z(\omega) = \omega S(\omega) \left(1 + \frac{i}{2Q} \right). \quad (43)$$

First, in Fig. 2 we show the effects of permeability on the tube wave slowness and attenuation in a mudcake-free borehole, for which $b=a$ and $W_{mc} \equiv 0$, and we compare against the results for an impermeable/elastic medium. Essentially these results had been presented previously by various authors (White, 1983; Cheng *et al.*, 1987; Schmitt *et al.*, 1988; Chang *et al.*, 1988; Winkler *et al.*, 1989; Norris, 1989). We see that the effect of a nonzero permeability is to increase the slowness of the mode and to increase the attenuation, especially so as the frequency is lowered. These effects are due to the ability of the borehole fluid to leak into the permeable rock and couple into the slow wave. These phenomena were first pointed out by White (1983). They were observed in the laboratory experiments and reported by Winkler *et al.* (1989), who showed that the predictions of the Biot theory were in excellent agreement with the data. At the limit of zero frequency, the open borehole tube wave slowness diverges and the value of Q^{-1} tends to 2. At high frequency, the specific attenuation decreases with increasing frequency.

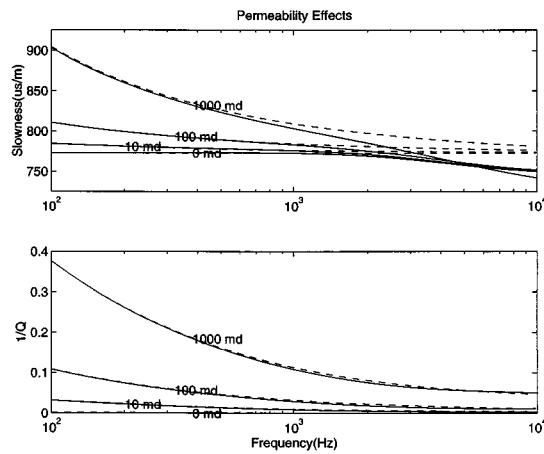


FIG. 2. Open borehole slowness and attenuation for various permeabilities. The membrane stiffness is zero as is the mudcake thickness. All other parameters as in Table I. The solid curves represent the results of the full numerical mode-search calculations while the dashed lines represent the low-frequency limit of Sec. III.

Furthermore, the open borehole tube wave slowness can become less than (faster than) the nonpermeable, elastic case. This phenomenon is exaggerated in a high-permeability situation and was easily observable in the laboratory experiments (Winkler *et al.*, 1989). Similar effects, where the slowness of the tube wave in a permeable rock at high frequency is less than that in an elastic formation, were pointed out by Norris (1989). Also shown in Fig. 2 are the results of the low-frequency analytic results discussed in the next section.

Next, Fig. 3 shows the tube wave characteristics for an open and for three sealed boreholes in a sample of 200 md permeability. The mudcake still has zero thickness ($a=b$) but has a finite membrane stiffness, $W_{mc} \neq 0$. Intuitively, this should act to reduce the fluid flow mechanism and hence the effects of permeability on tube wave characteristics. We see from Fig. 3 that this is certainly the case at low frequencies but that there is still very appreciable attenuation at higher

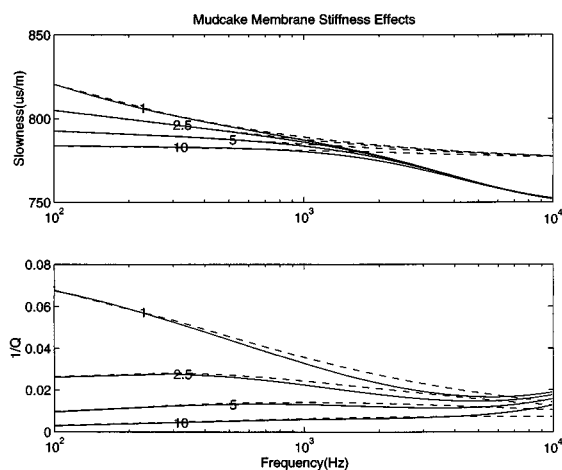


FIG. 3. Slowness and attenuation for varying mudcake membrane stiffnesses, in units of GPa/cm. The mudcake thickness is equal to zero and the formation permeability is equal to 200 mD. Same conventions as in Fig. 2.

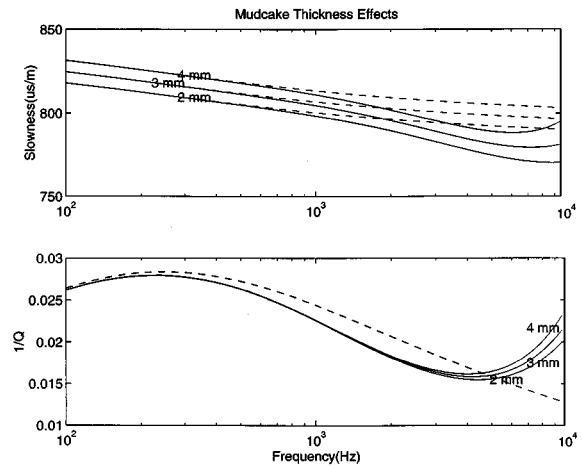


FIG. 4. Slowness and attenuation for varying mudcake thickness. The membrane stiffness is equal to 2.5 GPa/cm and the permeability is 200 mD. Other conventions as in Fig. 2.

frequencies. For membrane stiffnesses $W_{mc} \leq 1.0$ GPa/cm, the tube wave characteristics appear to be like an open borehole. Note that the slight increase in attenuation toward high-frequency begins to show in this scale. This attenuation effect is due to the generation of a slow wave through the solid. It is nonzero even if the fluid and the solid constituents are locked together, $U_{r2}(b^+) \equiv u_{r2}(b^+)$, at the boundary, which would be the case for $W_{mc} \rightarrow \infty$ as seen from Eq. (8). Our results are similar to those of Tang (1994).

As we mentioned earlier, our model neglects the permeability in the mudcake; this is justified because the permeability of a typical mudcake is on the order of 1 microdarcy (Ferguson and Klotz, 1954; Meeten and Sherwood, 1994). If taken literally, this value would imply $\kappa \approx 10^{11}$ gm cm⁻² s⁻¹ for a mudcake of thickness 1 mm, which in turn would imply $|W_{mc}| = |\omega \kappa| \approx 10^{5 \pm 1}$ GPa cm⁻¹, values so large that the borehole would always be effectively sealed. On the other hand, for reasonable values of the relevant parameters, the model of Tang (1994) easily gives values $W_{mc} \approx [1-10]$ GPa cm⁻¹, as we have it.

Finally, we investigate the effects of a finite mudcake thickness; as far as we are aware, such calculations have not been presented before. Even for a thin elastic layer with a low shear rigidity, the effect on tube wave slowness is quite significant. We use an elastic layer which has a compressional velocity of 1600 m/s, a shear velocity of 320 m/s and a density of 1.1 gm/cm³ as in Table I. The results are shown in Fig. 4 for thicknesses 2 mm, 3 mm, and 4 mm. The thickness effect is reduced if the shear rigidity increases; there is a simple explanation for this which we give in the next section. Roughly speaking, a finite layer thickness acts to change the slowness by a frequency independent shift, with little change in attenuation. These two effects also have a simple explanation, which we also present in the next section.

The values we have used for the mudcake parameters are best guesses. It is thought that the speed of a compressional wave in the mudcake is very similar to that in the

borehole mud itself (E. Fordham, private communication). It is known that the shear modulus in a gelled borehole mud is very small; if one extrapolates the data of Alderman *et al.* (1991) to a concentration of 50% one gets values $\mu \approx 10^7 - 10^8 \text{ dyn cm}^{-2}$ which presumably represent a lower bound to μ , as we have it. In an unpublished report, Tang and Martin (1994) also claim values of μ in the range $10^7 - 10^8 \text{ dyn cm}^{-2}$.

III. LOW-FREQUENCY LIMIT

The full model that we have presented so far depends upon 18 different parameters and it is not immediately obvious which parameters are the more important and which the less. Here, we develop a closed-form expression for $k_z(\omega)$ which is valid in the low-frequency limit. We follow closely the derivation of Chang *et al.* (1988) who derived the desired result in the specific case that there is no mudcake at all, $b=a$ and $W_{mc}=0$. They make the following assumptions: (i) The low-frequency limit of the Biot theory is valid; that is, the viscous skin depth is much larger than Λ . In this limit the fast wave and the shear wave have real-valued phase speeds, and the slow wave is diffusive $v_{c2}^2 = -i\omega C_D$, with a diffusion constant given by

$$C_D = \frac{\kappa_0}{\eta\phi^2} \frac{PR - Q^2}{P + R + 2Q} \quad (44)$$

and so the wave vector of the slow wave is

$$k_{c2}(\omega) \equiv \omega/v_{c2} = \sqrt{i\omega/C_D}. \quad (45)$$

(ii) The wavelengths of all propagatory modes are large compared with the relevant borehole dimensions. This means that the pertinent Bessel functions can be replaced by their small argument limits. This approximation is not applied to the slow wave, whose wave length may be small or large. (iii) Because of Eq. (45), the axial component of the slow wave's wave vector k_z is small compared to the radial. Thus

$$k_{rc2} = \sqrt{k_{c2}^2 - k_z^2} \approx k_{c2} \quad (46)$$

as long as $\omega \ll V_T^2/C_D$ where V_T is the tube wave speed. With these three approximations applied to the relevant 4×4 matrix, Chang *et al.* (1988), and independently Norris (1989), were able to derive an explicit closed-form solution:

$$k_z^2 = \omega^2 \rho_{f0} \left[\frac{1}{K_{f0}} + \frac{1}{N} + \frac{2}{bW_p} \right], \quad (47)$$

where

$$W_p(\omega) = - \frac{\eta C_D k_{c2} H_0^{(1)}(k_{c2}b)}{\kappa_0 H_1^{(1)}(k_{c2}b)} \quad (48)$$

is a frequency-dependent added stiffness due to the permeable borehole wall. This result corrected a slight inconsistency in an earlier result due to White (1983) in a quantitative manner.

We have attempted to derive the formula equivalent to Eq. (47) by direct application of these three approximations to our own 8×8 matrix. The resultant matrix is still so large that we were unable to use a symbolic manipulator; there are

approximately 10^4 terms. The problem is the elastic layer which we have effectively removed from the problem via the following procedure.

In the quasistatic limit, and within the elastic layer, we are justified in neglecting the z dependence of displacements and stresses with respect to their r dependence. In this limit the shear stress has a simple dependence upon r : $\tau_{rz1} = C/r$ for some constant C . Since $\tau_{rz1} = 0$ at the boundary with the fluid, it must vanish everywhere: $C=0$. In turn, this implies $u_{z1}(r) = 0$; the mudcake is in plane strain. The radial displacement and the radial stress on the inner surface of the layer can be related via elementary elasticity theory to their respective values on the outer surface:

$$\begin{bmatrix} u_{r1}(b^-) \\ \tau_{rr1}(b^-) \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} u_{r1}(a^+) \\ \tau_{rr1}(a^+) \end{bmatrix}. \quad (49)$$

We find

$$\begin{aligned} T_{11} &= \frac{b^2\mu + a^2(\lambda + \mu)}{(\lambda + 2\mu)ab}, & T_{12} &= \frac{b^2 - a^2}{2(\lambda + 2\mu)b}, \\ T_{21} &= \frac{2(\lambda + \mu)\mu(b^2 - a^2)}{(\lambda + 2\mu)ab^2}, & T_{22} &= \frac{a^2\mu + b^2(\lambda + \mu)}{(\lambda + 2\mu)b^2}. \end{aligned} \quad (50)$$

The eight boundary conditions, Eqs. (1)–(8) can be reduced immediately to six; Eq. (7) becomes $\tau_{rz2}(b^+) = 0$ and Eq. (5) is indeterminate. In the remaining equations $u_{r1}(b^-)$ and $\tau_{rr1}(b^-)$ can be related to $u_{r1}(a^+)$ and $\tau_{rr1}(a^+)$ via Eqs. (50) and these, in turn, can be related to the borehole fluid displacement and pressure via Eqs. (1) and (2). The result is a set of 4 boundary conditions:

$$\begin{aligned} T_{11}U_{r0}(a^-) - T_{12}P_0(a^-) \\ = (1 - \phi)u_{r2}(b^+) + \phi U_{r2}(b^+), \end{aligned} \quad (51)$$

$$T_{21}U_{r0}(a^-) - T_{22}P_0(a^-) = \tau_{rr2}(b^+) - \phi P_2(b^+), \quad (52)$$

$$0 = \tau_{rz2}(b^+), \quad (53)$$

$$\begin{aligned} T_{21}U_{r0}(a^-) - T_{22}P_0(a^-) \\ = -P_2(b^+) - W_{mc}\phi[U_{r2}(b^+) - u_{r2}(b^+)]. \end{aligned} \quad (54)$$

Following the previous procedure, we derive a matrix equation of the form $[M]\mathbf{A} = \mathbf{B}$ where $[M]$ is now a 4×4 matrix. As before, the condition for the existence of a nontrivial normal mode is that the determinant of that matrix be zero. To this point, the only approximation is that the thickness of the layer, $b-a$, is very small compared with the shear and compressional wavelengths in that layer. Therefore, the tube wave slowness and attenuation calculated from this smaller determinant should be fully equivalent to that calculated from the larger, subject to this restriction on the mudcake parameters.

Moreover, the smaller matrix lends itself to an analytic derivation of the low-frequency limit. We apply exactly the three conditions used by Chang *et al.* (1988) discussed above. We set the resultant determinant equal to zero and solve for $k_z(\omega)$. The result is

$$k_z^2(\omega) = \omega^2 \rho_{f0} \frac{(1/K_{f0}) + (bT_{22} + 2NT_{12})/(aNT_{11} + abT_{21}/2) + \{[aT_{21}/K_{f0} + 2T_{22}]N\}/\{[W_{mc} + W_p][aNT_{11} + abT_{21}/2]\}}{1 + \{aNT_{21}/[W_{mc} + W_p][aNT_{11} + abT_{21}/2]\}} \quad (55)$$

Equation (55) is the central result of this section. It is plotted by a dashed curve in Figs. 2–4 where we see that it does indeed agree well with the full numerical mode-search solution in solid curve, the better the agreement as the frequency is decreased. In the special case of Fig. 2 that there is no mudcake [$a \equiv b$ and $W_{mc} \equiv 0$], it reduces to the earlier result in Eq. (48). We note in passing that the slowness diverges and the specific attenuation tends to a value of 2, in the limit as the frequency goes to zero. In the special case of Fig. 3 that there is a membrane [$W_{mc} \neq 0$], but the mudcake thickness is zero [$a \equiv b$], it reduces to a result derived by Norris (1989). Here, the slowness tends to a finite value and the specific attenuation eventually drops to zero in the limit as the frequency goes to zero. In the special case that there is no permeability, $\kappa_0 \rightarrow 0$, $\Rightarrow k_{c2} \rightarrow \infty$, $\Rightarrow W_p \rightarrow \infty$, it reduces to a result which had also been derived by Norris (1990)

$$k_z^2(\omega) = \omega^2 \rho_{f0} \left[\frac{1}{K_{f0}} + \frac{1}{M_F} \right], \quad (56)$$

where

$$M_F = \frac{N[\lambda + 2\mu] + f_c(\lambda + \mu)(\mu - N)}{\lambda + 2\mu + f_c(N - \mu)}, \quad (57)$$

and $f_c = 1 - a^2/b^2$ is the area fraction of the borehole occupied by the mudcake.

In order to achieve some intuitive understanding of the effects of mudcake on tube wave characteristics, we further approximate Eq. (55). We note that the complicated looking denominator in Eq. (55) differs from unity by only a small part if *either* the mudcake thickness $b - a$ is small so that $T_{21} \approx 0$ or if permeability effects are small generally. When either is true the denominator may be expanded in a Taylor series: $(1 + x)^{-1} \approx 1 - x$. We find

$$k_z^2(\omega) = \omega^2 \rho_{f0} \left[\frac{1}{K_{f0}} + \frac{1}{M_F} + \frac{1}{[W_{mc} + W_p]} \frac{2}{b} \times \frac{(1 - f_c)N^2}{[N - f_c[(\lambda + \mu)/(\lambda + 2\mu)](N - \mu)]^2} \right]. \quad (58)$$

On an intuitive level, the effect of the mudcake is to change the slowness, but not the attenuation, of the tube wave through the term $1/M_F$, an effect which is present in non-permeable formations [viz., Eq. (55)]. The membrane, W_{mc} , acts as if it were an impedance in series with that due to fluid flow, W_p ; it reduces the effect of permeability on dispersion and on attenuation. There is a further modification by the factor

$$\frac{(1 - f_c)N^2}{[N - f_c[(\lambda + \mu)/(\lambda + 2\mu)](N - \mu)]^2};$$

which reflects the fact that the stress on the formation wall is different from its value in the borehole fluid. Oddly, this

factor can be greater than unity when $\mu \ll (\lambda, N)$.

Equations (55) and (58) are identical when the mudcake thickness is zero. There is a difference between them for the parameters used in Fig. 4, but it is small. The difference between them can become appreciable in the somewhat artificial case for which the mudcake thickness is finite, $T_{12} \neq 0$, but the membrane stiffness vanishes, $W_{mc} \equiv 0$. This guarantees that the approximation used in going from Eq. (55) to Eq. (58) is invalid because $\lim_{\omega \rightarrow 0} W_p = 0$.

Equation (58) gives a simple explanation of three observations from the previous section: The thickness dependence acts to add an (almost) constant shift to the slowness with little change in attenuation and the effect lessens if μ is increased. Indeed, the effect (almost) disappears if $\mu = N$, for which $M_F \equiv N$.

Equation (55) has been derived under the assumption that the shear and compressional wavelengths in the mudcake are large compared to $b - a$. When that is true, Eq. (55) is valid even when $\mu \ll \lambda$, as seems to be the case for realistic mudcakes. If μ is small enough, however, then at any finite frequency the shear wavelength can be *small* compared with $b - a$ and the previous derivation is invalid. In fact, Eq. (55) is wrong if one assumes $\mu \equiv 0$. In Fig. 5 we show the effect of varying the shear modulus of the mudcake, holding all other parameters including W_{mc} fixed. As μ is reduced, the low-frequency results tend to the limiting expression given by Eq. (55) with $\mu = 0$ but there is a shear wave resonance which moves down in frequency, as is apparent in Fig. 5. Eventually this, and other, shear wave resonances pass entirely through the frequency band indicated in the figures: The validity of Eq. (55) or (58) is restricted to ever lower frequencies.

It is a simple matter, however, to treat the case where the

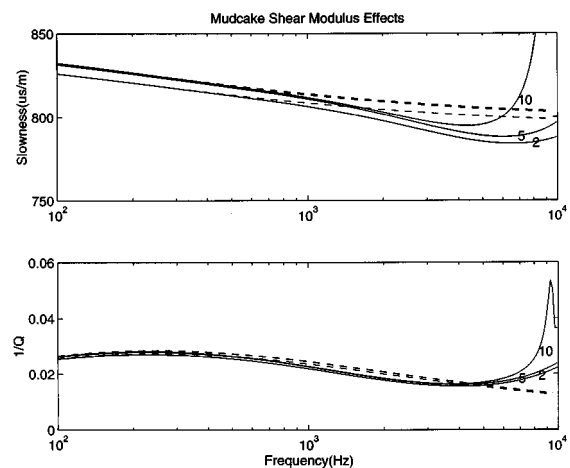


FIG. 5. Sensitivity of tube wave characteristics to variations in the mudcake shear modulus with mudcake thickness = 4 mm. The case $V_p/V_s = 5$ corresponds exactly to Fig. 4.

“mudcake” is a layer of fluid having a vanishing shear modulus. The result follows almost by inspection of Eq. (58). The region $r < b$ is occupied by an inhomogeneous fluid with an effective compressibility which is the volume weighted compressibility of the borehole fluid and of the “mudcake.” There is no solid layer. Therefore, when $\mu \equiv 0$ the low frequency limit is

$$k_z^2(\omega) = \omega^2[(1-f_c)\rho_{f0} + f_c\rho_{s1}] \left[\frac{1-f_c}{K_{f0}} + \frac{f_c}{\lambda} + \frac{1}{N} + \frac{1}{[W_{mc} + W_p]b} \right]. \quad (59)$$

That Eqs. (58) and (59) are different (even when $\mu = 0$ in the former) can be seen immediately. We note that these foregoing low-frequency results may predict inaccurate values of attenuation when the permeability is very large, for reasons spelled out by Johnson and Kostek (1995). In the range of parameters considered in the present article, this does not happen.

IV. CONCLUSIONS

We have presented a physically reasonable model for the effects of a mudcake on tube wave propagation in a porous and permeable formation. As in earlier work, the formation is treated within the context of the Biot Theory of acoustics in porous and permeable materials. The mudcake is modeled as an elastic layer of finite thickness. We have incorporated the assumption that there is a finite membrane stiffness at the mudcake-formation boundary. This membrane controls the extent to which the mudcake can flex into and out of the pore space of the formation and thus it controls coupling of the tube wave to the Biot slow compressional wave. It had previously been demonstrated that the effect of a finite permeability is to cause the slowness and the specific attenuation to increase with decreasing frequency, the more so the larger the permeability. In the present paper, our numerical calculations indicate that the effect of a finite value of the membrane stiffness acts to reduce the effects of permeability on both slowness and on attenuation. The effect of a finite thickness of mudcake is (mostly) a constant change in slowness with little change in attenuation. We have derived a closed-form expression for the low-frequency limit of the tube wave characteristics, which clarifies these observations. However, for a mudcake-sealed borehole there is an increase in attenuation toward high frequency due to finite coupling to the slow wave.

The mudcake model described here may be overly simplified considering the mud filtrate process and the details of mudcake growth. In the specific case that the mudcake is a nonwetting fluid, Nagy and Blaho (1994) suggest that the surface tension and the permeability itself may play a role in the sealing mechanism. In the case considered in the present article, Tang (1994) suggested that the mudcake shear rigidity and the pore size are the key factors in determining the membrane stiffness. However, to further define the mudcake dependence on either pore size, porosity, or permeability requires more empirical observations.

ACKNOWLEDGMENTS

We are grateful for many discussions with our colleagues Philip Cheung, Edmund Fordham, Benoit Froelich, Gerry Meeten, Tom Plona, Ken Winkler, and David Wilkinson.

- Alderman, N. J., Meeten, G. H., and Sherwood, J. D. (1991). “Vane rheometry of bentonite gels,” *J. Non-Newtonian Fluid Mech.* **39**, 291–310.
- Biot, M. A. (1956a). “Theory of propagation of elastic waves in a fluid-saturated porous solid, I. Low-frequency range,” *J. Acoust. Soc. Am.* **28**, 168–178.
- Biot, M. A. (1956b). “Theory of propagation of elastic waves in a fluid-saturated porous solid, II. High-frequency range,” *J. Acoust. Soc. Am.* **28**, 179–191.
- Chang, S. K., Liu, H.-L., and Johnson, D. L. (1988). “Low-frequency tube waves in permeable rocks,” *Geophysics* **53**, 519–527.
- Cheng, C. H., Jinzhong, Z., and Burns, D. R. (1987). “Effects of *in-situ* permeability on the propagation of Stoneley (tube) waves in a borehole,” *Geophysics* **52**, 1279–1289.
- Deresiewicz, H., and Skalak, R. (1963). “On uniqueness in dynamic poroelasticity,” *Bull. Seis. Soc. Am.* **53**, 783–788.
- Ferguson, C. K., and Klotz, J. A. (1954). “Filtration from Mud During Drilling,” *Petrol. Trans. AIME* **201**, 29–42.
- Fordham, E. (1996). Private communication.
- Johnson, D. L., Koplik, J., and Dashen, R. (1987). “Theory of dynamic permeability and tortuosity in fluid-saturated porous media,” *J. Fluid Mech.* **176**, 379–400.
- Johnson, D. L., Hemmick, D. L., and Kojima, H. (1994a). “Probing porous media with first and second sound. I. Dynamic permeability,” *J. Appl. Phys.* **76**, 104–114.
- Johnson, D. L., Plona, T. J., and Kojima, H. (1994b). “Probing porous media with first and second sound. II. Acoustic properties of water-saturated porous media,” *J. Appl. Phys.* **76**, 115–125.
- Johnson, D. L., and Kostek, S. (1995). “A limitation of the Biot-Gardner theory of extensional waves in fluid-saturated porous cylinders,” *J. Acoust. Soc. Am.* **97**, 741–744.
- Kurkjian, A. L., and Chang, S.-K. (1986). “Acoustic multipole sources in fluid-filled boreholes,” *Geophysics* **51**, 148–163.
- Meeten, G. H., and Sherwood, J. D. (1994). “The hydraulic permeability of bentonite suspensions with granular inclusions,” *Chem. Eng. Sci.* **49**, 3249–3256.
- Nagy, P. B., and Blaho, G. (1994). “Experimental measurements of surface stiffness on water-saturated porous solids,” *J. Acoust. Soc. Am.* **95**, 828–835.
- Norris, A. N. (1989). “Stoneley wave attenuation and dispersion in permeable formations,” *Geophysics* **54**, 330–341.
- Norris, A. N. (1990). “The speed of a tube wave,” *J. Acoust. Soc. Am.* **87**, 414–417.
- Rosenbaum, J. H. (1974). “Synthetic microseismograms: Logging in porous formations,” *Geophysics* **39**, 14–32.
- Schmitt, D. P. (1988). “Effects of radial layering when logging in saturated porous formations,” *J. Acoust. Soc. Am.* **84**, 2200–2214.
- Schmitt, D. P., Bouchon, M., and Bonnet, G. (1988). “Full-wave synthetic acoustic logs in radially semiinfinite porous media,” *Geophysics* **53**, 808–823.
- Tang, X. M. (1994). “Effects of mudcake on the measurement of fluid flow properties using borehole acoustic waves,” *SEG Expanded Abstracts*, 64th Annual Meeting, Los Angeles, p. 78, BG 3.4.
- Tang, X. M., Cheng, C. H., and Toksöz, M. N. (1991). “Dynamic permeability and borehole Stoneley waves: A simplified Biot–Rosenbaum model,” *J. Acoust. Soc. Am.* **90**, 1632–1646.
- Tang, X. M., and Martin, R. J. (1994). “High resolution evaluation of formation flow properties from a borehole acoustic imaging tool,” *Final Report, Gas Res. Inst. Contract 5093-260-2753* (unpublished).
- White, J. E. (1983). *Underground Sound-Application of Seismic Waves* (Elsevier, Amsterdam).
- Williams, D. M., Zemanek, J., Angona, F. A., Dennis, C. L., and Caldwell, R. L. (1984). “The long space acoustic logging tool,” *SPWLA, 25th Annual Logging Symposium*.
- Winkler, K. W., Liu, H.-L., and Johnson, D. L. (1989). “Permeability and borehole Stoneley waves: Comparison between experiment and theory,” *Geophysics* **54**, 66–75.

A boundary integral method for acoustic radiation and scattering

W. S. Hwang

Department of Naval Architecture and Ocean Engineering, National Taiwan University, Taiwan

(Received 18 October 1995; revised 15 July 1996; accepted 20 November 1996)

In this article, a computational method is proposed for computing the Helmholtz integral equation for acoustic radiation and scattering problems with a three-dimensional body shape. Unlike the previous boundary element method, the integration of acoustic properties is based on the global body surface instead of local elements. The approximation function of body geometry and the weighting function of numerical integration are independently chosen. When the singular kernels of Helmholtz equation are regularized as bounded, discontinuous functions, the order of integration polynomials for acoustic properties can be arbitrarily chosen, therefore, the accuracy and efficiency of computation can be increased. © 1997 Acoustical Society of America.

[S0001-4966(97)03505-4]

PACS numbers: 43.20.Rz, 43.20.Tb, 43.40.Rj [JEG]

INTRODUCTION

Boundary-integral methods have long been applied in acoustics for solving the solution of radiation and scattering problems. The most important advantage of this method is that the computational dimension of the problem can be reduced by one. For example, three-dimensional problems can be solved on the two-dimensional body surfaces. Besides, for exterior problems, the radiation condition is automatically incorporated in the integral formulation. However, the use of an integral-equation method accompanies the shortcoming of nonuniqueness, when applied to exterior problems at some characteristic values of the wave number. The characteristic wave numbers depend on the type of integral equation, the shape of bodies, and the boundary conditions of body surfaces.

To overcome the nonuniqueness problem, several methods have been proposed in the literature. Schenck¹ proposes additional constraints of the Helmholtz integral representation inside the boundary. He suggests a method of overdetermining the system with additional algebraic equations by a combined Helmholtz integral equation formulation (CHIEF). The interior points used in the CHIEF representation are called CHIEF points. The overdetermining algebraic equations may be solved by a least-squares procedure. Seybert *et al.*²⁻⁴ and Seybert and Rengarajan⁵ use the CHIEF method and function approximation technique on the surface, well-developed in finite element methods, to model acoustic radiation and scattering problems successfully. However, when CHIEF points fall on the nodal surfaces of the corresponding interior problem, the additional equations by CHIEF points fail to provide linearly independent constraints. Therefore, modified versions of CHIEF formulation are proposed. Enhanced CHIEF methods that incorporated the first-order and the second-order derivatives of the interior Helmholtz equations are presented.⁶⁻⁹ Although these methods do not solve the nonuniqueness problem completely, they do improve the original CHIEF method.

Another well-known formulation to overcome the nonuniqueness problem is proposed by Burton and Miller.¹⁰ This

approach combines the surface Helmholtz integral equation and its associated normal derivative equation. Although the nonuniqueness problem is solved for all characteristic wave numbers, the numerical implementation of the hypersingular integral is much more difficult than that of the CHIEF method. Many researchers¹¹⁻²¹ transform the hypersingular integral to reduce the singularity order of kernel and improve the scheme of numerical computation to reduce the execution time. Cunefare and Koopmann²² slightly vary the method of Burton and Miller by setting the field points of the kernels of the integrals into interior domain instead of boundaries. The complexity of hypersingular kernel is avoided from their computation.

The comparison of the CHIEF method and the Burton and Miller approach is well done in the literature. The advantage of CHIEF is easier to implement and well-suited to low-frequency application, but it becomes increasingly difficult to apply when the frequency increases. The approach of Burton and Miller becomes more efficient as the wave number increases. The use of the CHIEF method is very popular in the industry and science because of the easy numerical implementation, and researchers put more attention on the nonuniqueness problem instead of efficiency. For example, the quadrilateral elements are widely used on the body surface for the approximation of geometry but it is not really a very efficient method for numerical integration by using low-order integration rules only in the local element. The accuracy of numerical integration can be increased by using high-order quadrature formula with the same number of integration points. Therefore, the rearrangement of integration points based on the global surface instead of local elements is a much more efficient way. Even the information of geometry may still be supplied from local elements.

In order to use high-order quadrature formula, the singular points in the integral equation must be removed or regularized. Recently, Huang and Chwang²³ removed the singular kernel of Helmholtz integral equation by an analytic procedure and directly solved the integral equation without

applying any approximation functions of boundary element methods. However, Huang and Chwang's result contained an additional volume integral in their integral formulation. In this paper, the focus is put on the promotion of efficiency of solving the Helmholtz integral equation, therefore, Gauss's flux theorem and property of equipotential surface will be applied to remove the kernel singularity of the surface Helmholtz integral equation without additional volume integrals. Then arbitrary-order Gaussian quadrature formula can be applied in the surface integral directly.

I. THEORY AND FORMULATION

The standard three-dimensional Helmholtz integral formula for velocity potential in an external problem is

$$\epsilon \Phi_S(P) = \int \int_S \left[\Phi_S(Q) \frac{\partial G}{\partial n_Q}(P, Q) - G(P, Q) \frac{\partial \Phi_S(Q)}{\partial n_Q} \right] dS_Q, \quad (1)$$

where P denotes the field point and Q denotes the source point on the surface of the body. In Eq. (1), Φ_S is the scattering velocity potential satisfying Sommerfeld radiation condition and Helmholtz equation $\nabla^2 \Phi_S + k^2 \Phi_S = 0$ for time-harmonic waves $G(P, Q)$ is the free-space Green's function of the Helmholtz equation,

$$G(P, Q) = \frac{e^{ikr}}{r(P, Q)}, \quad (2)$$

where k is the wave number and r the distance between $P(x, y, z)$ and $Q(\xi, \eta, \zeta)$. The coefficient ϵ has the value 0 for P inside surface S , the value 4π for P outside surface S , and the value 4π minus the solid angle of point P for P on surface S . For simplicity, a point P on the region of smooth surface is considered here only, hence ϵ has value 2π for point P on surface S .

The total velocity potential is the summation of velocity potentials of incoming and scattering waves such that

$$\Phi = \Phi_0 + \Phi_S, \quad (3)$$

where Φ_0 denotes the velocity potential of incoming waves. When point P is located outside surface S , the total velocity potential can be expressed as

$$\Phi(P) = \frac{1}{4\pi} \int \int_S \left[\Phi(Q) \frac{\partial G}{\partial n_Q}(P, Q) - G(P, Q) \frac{\partial \Phi(Q)}{\partial n_Q} \right] dS_Q + \Phi_0(P). \quad (4)$$

When point P lies on the smooth part of surface S , the value of ϵ is 2π . The Helmholtz integral equation can be modified as

$$\Phi(P) = \frac{1}{2\pi} \int \int_S \left[\Phi(Q) \frac{\partial G}{\partial n_Q}(P, Q) - G(P, Q) \frac{\partial \Phi(Q)}{\partial n_Q} \right] dS_Q + 2\Phi_0(P). \quad (5)$$

In the present study, a hard body is considered, $\partial \Phi / \partial n = 0$, on surface S , so the second term of the integrand vanishes. In order to solve Eq. (5) efficiently, an intermediate step is needed. The basic idea was developed in the hydrodynamics.²⁴ Since the kernel function becomes singular when points P and Q are coincident, the direct numerical integration will become less accurate. By subtracting a different singular kernel from the integrand of Eq. (5), whose integral result is known analytically, and adding back its result of integration, the total velocity potential becomes easier for subsequent numerical calculation:

$$\Phi(P) = \frac{1}{4\pi} \int \int_S \left[\Phi(Q) \frac{\partial G}{\partial n_Q}(P, Q) - \Phi(P) \frac{\partial}{\partial n_Q} \left(\frac{1}{r}(P, Q) \right) \right] dS_Q + \Phi_0(P), \quad (6)$$

where the surface integration of $-(\partial/\partial n)(1/r)$ is the solid angle in geometry. In order to understand the behavior of the kernel of Eq. (6), when point Q approaches P , the body surface is assumed to be composed of regular surface elements, $f(x, y, z) = 0$, in the sense of Kellogg,²⁵ with twice continuous differentiability. The direction cosines at the point Q of surface are given by $(f_\xi, f_\eta, f_\zeta)/D$, where D is the normalized length. The kernel of integral Eq. (6) can be expanded as

$$\frac{1}{r^3 D} (\mathbf{r} \cdot \nabla_\xi f) [\Phi(Q) e^{ikr}(ikr - 1) + \Phi(P)], \quad (7)$$

where $\mathbf{r} = (\xi - x, \eta - y, \zeta - z)$. When point Q is close to point P , we can write the Taylor expansion,

$$f(x, y, z) = f(\xi, \eta, \zeta) - \mathbf{r} \cdot \nabla_\xi f + \frac{1}{2!} [(x_1 - \xi)^2 f_{\xi\xi} + \dots]. \quad (8)$$

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

$$\mathbf{r} \cdot \nabla_\xi f \approx O(r^2). \quad (9)$$

When Q approaches P , the bracket term in (7) can also be expanded, if the first derivatives of $\Phi(Q)$ are continuous:

$$\Phi(Q) e^{ikr}(ikr - 1) + \Phi(P) = \mathbf{r} \cdot \nabla_x \Phi(P) + O(r^{1+\alpha}), \quad (10)$$

where $\alpha > 0$. Substituting the results in (9) and (10) into (7), we observe that the leading order in $(\xi - x), (\eta - y), (\zeta - z)$ is the third order in the numerator, and the denominator of (7) is also of the third order of r , so the ratio is indeterminate as Q approaches P . Although the kernel of (6) is discontinuous when P and Q are coincident, it is antisymmetric and bounded when Q is close to P . The integral of (7) over a small area symmetric about P approaches zero, therefore, we set the integrand of Eq. (6) equal to zero when Q is coincident with P in an average sense. Then, the numerical integration in Eq. (6) can be easily computed by any numerical quadratural formulation. The above statement is true when P and Q are on the region of smooth surfaces of twice continuous differentiability. When the body surface includes nonsmooth edges, the kernel of Eq. (6) cannot be set to zero on those points of edges. Actually, it depends on the solid angle and the first derivative of velocity potential at that point. Therefore, the body surface is broken into pieces of

regular surfaces and the total surface integral has to be summed up by the integral of all regular surfaces. In each regular surface element, collocation points are chosen to be Gaussian points, since the Gaussian points do not locate at the edges of regular surfaces. Then, the kernel of Eq. (6) can still be set to zero when P and Q are coincident, and the formula of Gaussian quadrature can be applied.

When radiation problem is considered, $\Phi_0(P)$ in Eq. (6) can be simply replaced by

$$\Phi_0(P) = -\frac{1}{4\pi} \int \int_S G \frac{\partial \Phi}{\partial n_Q} dS_Q. \quad (11)$$

For radiation problem, this is an another kind of singular integral kernel, $1/r$, and usually this singular integral can be removed by using the transformation of polar coordinates on the local elements. Here, we introduce a different approach to solve this kind of singular integral without such a transformation and do not increase too much computation. In potential theory, we can find a source distribution, $S(Q)$, on body surface which makes the surface an equipotential of potential Φ_0 . This relation is also called electrostatic capacitance in electrostatics. The source distribution satisfies a homogeneous integral equation and the detailed derivation can be found in Jaswon and Symm:²⁶

$$\frac{1}{2\pi} \int \int_S S(Q) \frac{\partial}{\partial n_P} \left(\frac{1}{r} \right) dS_Q = -S(P). \quad (12)$$

In fact, $S(P)$ is an eigenfunction of Eq. (12). The source distribution cannot be determined uniquely, and we just choose one solution since only the relative values and its corresponding potential value are concerned. Before directly solving Eq. (12), we subtract a known singular integral from Eq. (12) and add its result back to remove the singularity of integrand based on the same argument as Eq. (6), and express it in an iterative form.

$$S_{n+1}(P) = S_n(P) - \frac{1}{2\pi} \int \int_S \left[S_n(Q) \frac{\partial}{\partial n_P} \left(\frac{1}{r} \right) - S_n(P) \frac{\partial}{\partial n_Q} \left(\frac{1}{r} \right) \right] dS_Q. \quad (13)$$

Those coefficients required for solving $S(P)$ in Eq. (13) actually are already obtained in the computation procedure for constructing the matrix in Eq. (6), therefore, no additional computation effort is necessary to form the matrix for Eq. (13). After $S(P)$ is solved, the equipotential Φ_e can be evaluated by the equation

$$\Phi_e + \int \int_S \frac{S(Q)}{r} dS_Q = 0. \quad (14)$$

Since Φ_e is proven to be constant in the whole domain inside the body surface in Jaswon and Symm,²⁶ for simplicity, we just choose one field point inside the body to get the equipotential value. The right-hand side term in Eq. (11) can finally be computed by subtracting Eq. (14) times $(\partial \Phi / \partial n_P) \times [1/S(P)]$.

$$\begin{aligned} \int \int_S G \frac{\partial \Phi}{\partial n_Q} dS_Q &= \int \int_S \left[G \frac{\partial \Phi(Q)}{\partial n_Q} - \frac{1}{r} \frac{S(Q)}{S(P)} \frac{\partial \Phi(P)}{\partial n_P} \right] \\ &\quad \times dS_Q - \frac{\Phi_e}{S(P)} \frac{\partial \Phi(P)}{\partial n_P}. \end{aligned} \quad (15)$$

Again, the singular kernel of type $1/r$ is removed in the bracket on the right-hand side of Eq. (15) by the similar argument as Eq. (6), therefore, the integration can be implemented directly.

At the first glance, it seems much more complicated to evaluate Eq. (11) by Eq. (15) than by local polar transformation. In fact, it is quite efficient since the source density $S(P)$ is dependent on the geometry only and nothing to do with the frequency, therefore, the solution of Eq. (13) is smooth and easy to be computed. Once Eq. (13) is solved, it can be repeatedly used for every different frequency, while the local polar transformation must be repeated for every different frequency.

II. NUMERICAL EXAMPLES

In the traditional boundary element methods, the surface is discretized into a number of elements. In each element, the acoustic variables are approximated by a given function, say, linear or quadratic polynomials, in which the technique is fully developed in the finite element methods. The surface integral is summed up by the quadrature of each local element, in which lower-order approximated functions are used. In the present study, the surface integral can be integrated directly and globally, since the singular behavior is removed by an analytical formula, and the accuracy of numerical integration is increased by using high-order Gaussian quadrature formula.

In this section, a grid mesh is generated on the surface S and it depends on the numerical quadrature formula being chosen. The computing points are named either i or j counting from one to n total grid points. Equation (6) is discretized as

$$\Phi_i = \frac{1}{4\pi} \sum_{j=1, j \neq i}^n w_j (\Phi_j M_{ij} - \Phi_i R_{ij}) + \Phi_{0i}, \quad i = 1, \dots, n, \quad (16)$$

$$M_{ij} = \frac{\partial G}{\partial n_Q} (P, Q), \quad (17)$$

$$R_{ij} = \frac{\partial}{\partial n_Q} \left(\frac{1}{r} (P, Q) \right), \quad (18)$$

where Φ_i represents $\Phi(P)$ at mesh points, Φ_j represents $\Phi(Q)$, Φ_{0i} represents $\Phi_0(P)$, and w_j is the weighting factor for the surface integral. The evaluation of R_{ij} actually a by-product from M_{ij} , does not increase any amount of computer time. Another advantage of Eq. (12) is the flexibility. If the body surface has symmetric or axisymmetric property in geometry and boundary conditions, Eq. (12) can directly utilize those properties without reformulation. For instance, the

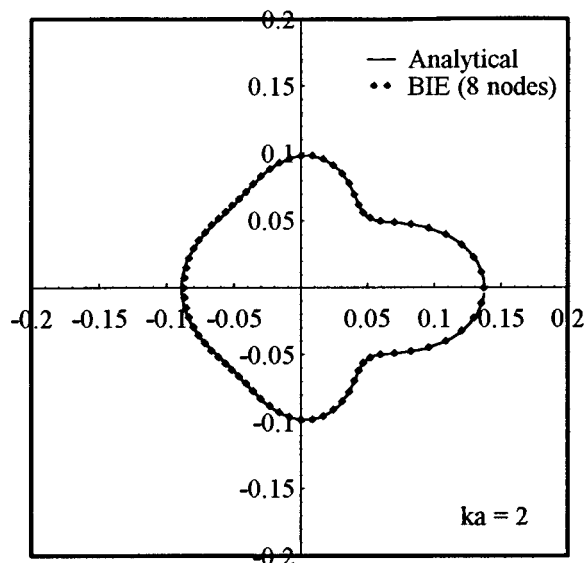


FIG. 1. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka=2$ and $r=5a$.

work by Seybert *et al.*⁴ is a special representation for axisymmetric cases, and this expression is equivalent to Eq. (16) if the axisymmetric property is considered.

The weighting factors should be calculated from the definition of body geometry, however, in general, the exact values are not easy to obtain directly for an irregular surface. The generation of mathematical description of geometry is necessary for the computation purpose. In the conventional boundary element method, the quadratic polynomials are most popularly used, but the accuracy of the quadratic approximation, usually, is not very good if elements are large. In the present method, the position of body surface, its normal vector, and the weighing factor of the surface integral are not limited to be determined from quadratic polynomials. Any approximated functions, such as *B*-spline or rational functions, which can generate the desired geometric data, are available for the present method.

Unlike the conventional method, no shape functions are assumed in the present method. All the unknowns are located directly at the integration points, therefore, the number of collocation points is always the same as the number of integration points no matter how one chooses the numerical integration rule. When Gaussian quadrature is applied, the number of equations is exactly the number of Gaussian points.

A. Application to scattering

The incoming unit plane wave travels to the right along the x axis and is described as $\Phi_0 = \exp(ikx)$. For the scattering wave from a sphere of radius a the scattered velocity potential at a distance r from the center of sphere and an angle θ from the x axis is given by

$$\Phi_S(r, \theta) = \sum_{n=1}^{\infty} -\frac{i^n(2n+1)j'_n(ka)}{h'_n(ka)} P_n(\cos \theta) h_n(kr), \quad (19)$$

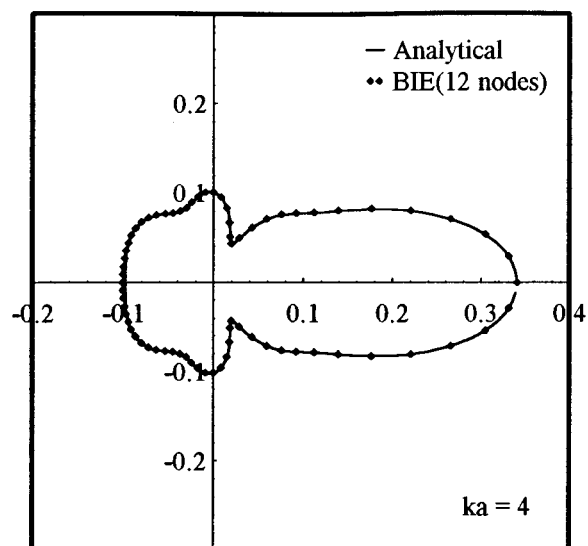


FIG. 2. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka=4$ and $r=5a$.

where P_n is the Legendre function of the first kind, h_n the spherical Hankel function of the first kind, and j_n the spherical Bessel function of the first kind. Figures 1–3 show the plots of real and imaginary parts of Φ_s versus polar angle of the rigid sphere on the sphere surface, $r=5a$, when nondimensional wave number $ka=2, 4$, and 8 . The analytical results are plotted as solid lines for absolute values of Φ_s , and the numerical results are marked by symbol spots for comparison. Figures 4 and 5 show the scattered velocity potential on the sphere surface. Due to an axisymmetric capability embedded in the computing code, only a semicircle is computed.

As shown in Fig. 1, $ka=2$, an eight-point Gaussian formulation has enough accuracy in the computation, while the nine nodes are used to obtain the similar result in Seybert

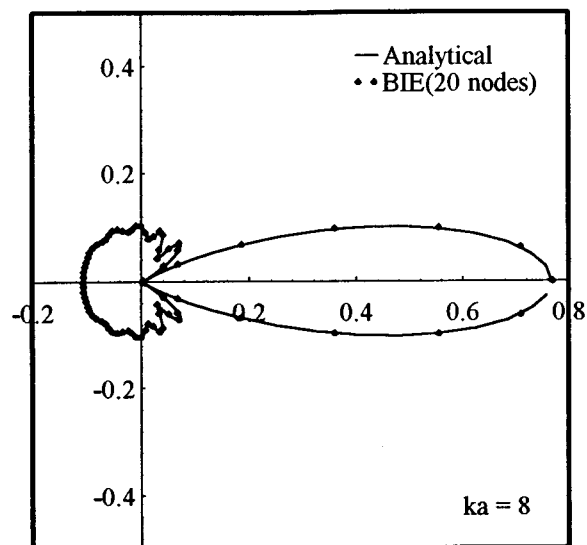


FIG. 3. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka=8$ and $r=5a$.

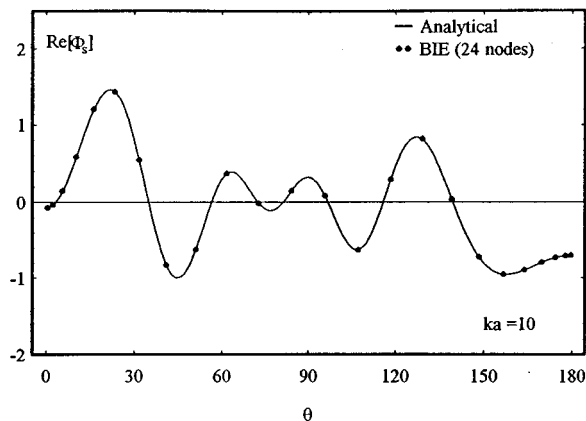


FIG. 4. Real part of ϕ_s on the surface of a rigid sphere when $ka=10$.

*et al.*⁴ by quadratic polynomial elements. Hence, there is not much difference between the present and the conventional methods in the low-frequency range. However, when the frequency increases, for $ka=4$, Seybert *et al.*⁴ used a 17-node discretization to maintain the required accuracy, while the present method only needs a 12-point Gaussian discretization, as shown in Fig. 2. It saves about 30% number of nodes. For higher frequency, $ka=8$, a 20-point discretization is satisfied as shown in Fig. 3 for the present method. Though Seybert *et al.* did not show this case, we expect that the difference of node number will enlarge between two methods, since the number of required nodes for the conventional method is about linearly proportional to the frequency. When the high-order quadrature formula is applied, the rate of required node increment grows slower as we can see in Figs. 1–3. Therefore, the present method becomes more efficient than the conventional methods when the frequency increases.

In Figs. 4–5, 24 collocation nodes are distributed on a semicircle according to the 24-point Gaussian formula. Both the scattering pattern and surface potential match the analytical results very well.

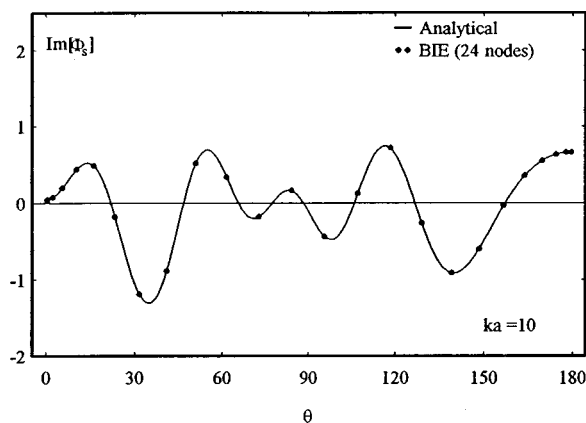


FIG. 5. Imaginary part of ϕ_s on the surface of a rigid sphere when $ka=10$.

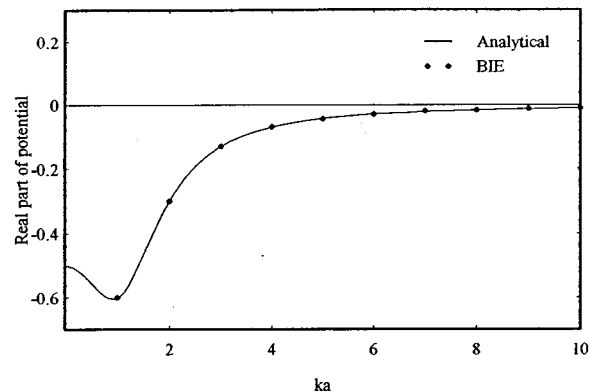


FIG. 6. Real part of ϕ/a on the surface of an oscillating sphere.

B. Application to radiation

An oscillating sphere is considered here. The analytical solution of velocity potential for the oscillating sphere of radius a with a radial velocity $\cos \theta$ is given by

$$\Phi(P) = \left(\frac{a}{r}\right)^2 \cos \theta \frac{a(1-ikr)(k^2a^2 - 2 - 2ika)}{k^4a^4 + 4} \times e^{ik(r-a)}. \quad (20)$$

In Figs. 6 and 7, the real and imaginary parts of the nondimensional surface velocity potential are plotted against the nondimensional wave number ka at $\theta=0.2$ rad the first Gaussian point in the 24-point Gaussian formula. Actually it is not necessary to use 24 computing points in the low-frequency range, but for the easy comparison we adopt the same computing points in different frequency. One CHIEF point is located in the position (0.5, 0.5, 0.5) to avoid the irregular frequency. Again, an excellent result is obtained.

III. CONCLUSIONS

It has been demonstrated that we can use the present method to regularize the weakly singular kernels of the Helmholtz integral equation in three-dimensional problems. This method can also be applied to the two-dimensional Helmholtz integral equation with a little modification of the regularized kernel function, although we do not show the details here.

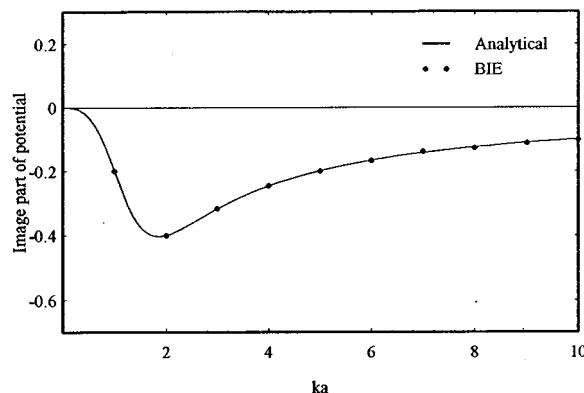


FIG. 7. Imaginary part of ϕ/a on the surface of an oscillating sphere.

It is very neat and efficient that the computation of the singular integral can be directly accomplished by Gaussian quadrature without special formulation near singular regions and the accuracy of computation is promoted by using high-order Gaussian quadrature without extra work. The computing time is reduced especially in the middle and high-frequency range because the required collocation points are decreased. When Gaussian points are used as collocation points, those calculation points will not locate at the nonsmooth edges or corners, where the normal vectors do not exist. Therefore, no especial treatment is needed for those particular points and it is convenient for programming. This idea can be applied to hypersingular integral equations also.

ACKNOWLEDGMENTS

The work in this paper was supported by the National Science Council, R.O.C., under Grant No. NSC 81-0209-E-002-07.

- ¹H. A. Schenck, "Improved integral formulation for acoustic radiation problems," *J. Acoust. Soc. Am.* **44**, 41–58 (1968).
- ²A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "Application of the BIE method to sound radiation problem using an isoparametric element," *ASME Trans. J. Vib. Acoust. Stress Reliabil. Des.* **106**, 414–420 (1984).
- ³A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "An advanced computational method for radiation and scattering of acoustic waves in three dimensions," *J. Acoust. Soc. Am.* **77**, 362–368 (1985).
- ⁴A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "A special integral equation formulation for acoustic radiation and scattering for axisymmetric bodies and boundary conditions," *J. Acoust. Soc. Am.* **80**, 1241–1247 (1986).
- ⁵A. F. Seybert and T. K. Rengarajan, "The use of CHIEF to obtain unique solution for acoustic radiation using boundary integral equations," *J. Acoust. Soc. Am.* **81**, 1299–1306 (1987).
- ⁶J. Zhou and H.-Z. Wang, "Calculating acoustic radiation from closed bodies using BIE Method," *Proceedings of the 2nd Marine and Weapon Underwater Noise Meeting of China*, Yangzhou, 1987.
- ⁷T. W. Wu and A. F. Seybert, "Acoustic radiation and scattering," in *Advanced Boundary Element Methods in Acoustics*, edited by R. D. Ciskowski and C. A. Brebia (Computational Mechanics, Southampton, U.K., 1991), Chap. 3.
- ⁸T. W. Wu and A. F. Seybert, "A weighted residual formulation for the CHIEF methods in acoustics," *J. Acoust. Soc. Am.* **90**, 1608–1614 (1991).
- ⁹D. J. Segalman and D. W. Lobitz, "A method to overcome computational difficulties in the exterior acoustics problem," *J. Acoust. Soc. Am.* **91**, 1855–1861 (1992).
- ¹⁰A. J. Burton and G. F. Miller, "The application of integral equation methods to the solution of some exterior boundary-value problems," *Proc. R. Soc. London, Ser. A* **323**, 201–210 (1971).
- ¹¹W. L. Meyer, W. A. Bell, B. T. Zinn, and M. P. Stallybrass, "Boundary integral solutions of three dimensional acoustic radiation problems," *J. Sound Vib.* **59**, 245–262 (1978).
- ¹²W. L. Meyer, W. A. Bell, M. P. Stallybrass, and B. T. Zinn, "Prediction of the sound field radiated from axisymmetric surfaces," *J. Acoust. Soc. Am.* **65**, 631–638 (1979).
- ¹³T. Terai, "On calculation of sound fields around three dimensional objects by integral equation methods," *J. Sound Vib.* **69**, 71–100 (1980).
- ¹⁴I. C. Mathews, "Numerical techniques for three-dimensional steady-state fluid-structure interaction," *J. Acoust. Soc. Am.* **79**, 1317–1325 (1986).
- ¹⁵S. Amini and D. T. Wilton, "An investigation of boundary element methods for the exterior acoustic problem," *Comput. Methods Appl. Mech. Eng.* **54**, 49–65 (1986).
- ¹⁶Z. Reut, "On the boundary integral methods for the exterior acoustic radiation problem," *J. Sound Vib.* **103**, 297–298 (1985).
- ¹⁷C. C. Chien, J. Rajiyah, and S. N. Atluri, "An effective method for solving the hypersingular integral equations in 3-D acoustics," *J. Acoust. Soc. Am.* **88**, 918–937 (1990).
- ¹⁸G. Krishnasamy, L. W. Schmerr, T. J. Rudolph, and F. J. Rizzo, "Hypersingular boundary integral equations: some applications in acoustic and elastic wave scattering," *J. Appl. Mech.* **57**, 404–414 (1990).
- ¹⁹T. W. Wu, A. F. Seybert, and G. C. Wan, "On the numerical implementation of a Cauchy principal value integral to insure a unique solution for acoustic radiation and scattering," *J. Acoust. Soc. Am.* **90**, 554–560 (1991).
- ²⁰T. W. Wu and G. C. Wan, "Numerical modeling of acoustic radiation and scattering from thin bodies using a Cauchy principal integral equation," *J. Acoust. Soc. Am.* **92**, 2900–2906 (1992).
- ²¹Yijun Liu and F. J. Rizzo, "A weakly singular form of the hypersingular boundary integral equation applied to 3-D acoustic wave problems," *Comput. Methods Appl. Mech. Eng.* **96**, 271–287 (1992).
- ²²K. A. Cunefare and G. Koopmann, "A boundary element method for acoustic radiation valid for all wave numbers," *J. Acoust. Soc. Am.* **85**, 39–48 (1988).
- ²³C. J. Huang and A. T. Chwang, "Diffraction of acoustic waves by rigid plane baffles," *J. Acoust. Soc. Am.* **95**, 668–680 (1994).
- ²⁴L. Landweber and M. Macagno, "Irrotational flow about ship forms," Iowa Institute of Hydraulic Research, The University of Iowa, IIHR Report No. 123, 1969.
- ²⁵O. D. Kellogg, *Foundations of Potential Theory* (Dover, New York, 1954).
- ²⁶M. A. Jaswon and G. T. Symm, *Integral Equation Methods in Potential Theory and Elastostatics* (Academic, London, 1977).

Hypersingular boundary integral equations for exterior acoustic problems

W. S. Hwang

Department of Naval Architecture and Ocean Engineering, National Taiwan University, Taiwan

(Received 29 November 1995; revised 15 October 1996; accepted 20 November 1996)

A bounded form of the hypersingular Helmholtz integral equation for 3-D acoustic problems is developed in this paper. All integrals with singular kernels are regularized within each regular surface element, in the sense of Kellogg, with twice continuous differentiability. The standard Gaussian quadrature formula is applied without any special treatment for all the computational points. The collocation points are chosen to be the Gauss–Legendre nodes and no interpolation function is assumed for acoustic variables. © 1997 Acoustical Society of America. [S0001-4966(97)03605-9]

PACS numbers: 43.20.Rz, 43.20.Tb, 43.40.Rj [JEG]

INTRODUCTION

The boundary integral method has been used for a long time in solving radiation and scattering problems in many scientific fields, such as potential theory, elastostatics, and acoustics. For exterior problems, the boundary integral equation accompanies the nonuniqueness problem at some characteristic wave numbers, which are corresponding to interior problems. However, the nonuniqueness is purely a mathematical drawback and does not have any physical meaning. To remove the fictitious eigenfrequency difficulty from exterior problems is one of the major tasks in the research of boundary integral equations.

A combined Helmholtz integral equation formulation (CHIEF), proposed by Schenck,¹ is a popular and simple method to circumvent the nonuniqueness problem. In this method, a few additional equations in the interior domain are added to the system of equations for the boundary Helmholtz integral equations and the overdetermining system is solved by a least-square procedure. However, the criterion how to choose the additional points inside the domain is, somehow, ambiguous, since those additional equations from CHIEF points may not provide enough linearly independent constraints, if the CHIEF points fall on nodal surfaces of the corresponding interior problems. Some modified CHIEF methods are proposed by several other researchers,^{2–6} however, they still do not solve the nonuniqueness problem completely, especially, in high-frequency range.

Another famous formulation to overcome the nonuniqueness problem at characteristic frequencies was proposed by Burton and Miller.⁷ In their paper, it is proved that the linear combination of the Helmholtz integral equation and its normal derivative can provide unique solutions at all frequencies, if the imaginary part of the coupling factor is nonzero. A different kind of difficulty from their formulation is that the normal derivative of the Helmholtz integral equation contains a hypersingular integral.

In Burton and Miller's original paper, two methods are proposed to handle the hypersingular kernel. One is to transform the hypersingular kernel into tangential derivatives on the surface and the other is to employ a regularization method to reduce the order of singularity by double surface

integrals. The double surface integral method is inefficient from a computational point of view and, therefore, less reported in the literature. One of these results, by Amini and Wilton,⁸ used bicubic *B* splines for the approximation of surface and Lagrangian interpolations for the approximation of solutions. On the other hand, the tangential derivative regularization is utilized by most researchers. Meyer *et al.*⁹ proposed one tangent derivative formula to circumvent the hypersingularity in which the body surface was simulated by flat elements and constant values of variables were assumed over each element. Mathews¹⁰ further improved this regularization algorithm by quadratic isoparametric surface elements. However, such approaches result $1/r$ type singular integrals even after a polar coordinate transformation. Special quadrature schemes are needed to accurately evaluate these singular integrals, but they are expensive in general. Terai¹¹ transformed the hypersingular integral into a contour integral with the aid of some analytical analysis, but this approach was limited in the application to flat elements with constant values of variables over each element. More detailed review can be found in Chien *et al.*¹²

More recently, solutions of potential theory are utilized to simplify the calculation of the hypersingular kernel from the Helmholtz integral equation by several researchers. Some authors^{13,14} consider the hypersingular integral equations as finite-part integrals based on the Hadamard sense, and employ certain identities from the Laplace equation and Stokes theorem to reduce the orders of hypersingular integrals into a weakly singular form. Another approach, developed by Chien *et al.*,¹² is to regularize the hypersingular integrals in the Helmholtz integral equations by the results from an associated Laplace equation in the interior domain, and it turns out a Cauchy principal value type singularity in the integral equation after the regularization. However, the adaptive subdivision around singular points is still needed in the numerical computation. In the meantime, Wu and his co-workers^{15,16} apply the tangential derivative method, originally derived by Maue,¹⁷ for the normal derivative integral equation. This regularized normal derivative integral equation converges in the Cauchy principal value sense rather than in the Hadamard finite-part sense. However, the number

of collocation points in this method is always greater than the number of nodal points, therefore, a least-square procedure is necessary for solving such an overdetermined system of equations.

In a recent paper,¹⁸ the author regularizes the singular kernels of the conventional Helmholtz integral equation by some identities from Laplace equations. In the conventional integral equation, it contains two kinds of weakly singular kernels. One of them, the normal derivative of Green's function, is frequently interpreted as $1/r^2$ type of singularity (Cauchy principal value singularity) in 3-D, but actually it is weakly singular only. The author reformulated the original weakly singular integral equation into a bounded form and solved it by an arbitrary-order boundary element method. In the present study, an advanced scheme is used to regularize the hypersingular kernel ($1/r^3$ type singularity in 3-D) of the normal derivative of the Helmholtz integral equation by using properties from associated Laplace equations and it turns out that all the singular kernels in the Helmholtz integral equation and its normal derivative are regularized as bounded discontinuous functions. Therefore, any order of Gaussian quadrature formula can be directly applied to solve the problem without any specially numerical treatment in the present paper.

I. FORMULATION OF INTEGRAL EQUATION

A. Helmholtz integral equation

The Helmholtz integral formula for the exterior acoustic problem can be written as

$$\bar{\epsilon}\Phi(P) = \frac{1}{4\pi} \int_S \left[\Phi(Q) \frac{\partial G}{\partial n_Q}(P, Q) - G(P, Q) \frac{\partial \Phi(Q)}{\partial n_Q} \right] dS_Q + \Phi_0(P), \quad (1)$$

where Φ is the scalar velocity potential satisfying the homogeneous Helmholtz equation, P the field point, Q the source point on the surface of body, n_Q the outward directed normal from surface S , and Φ_{in} the velocity potential of incident wave. The coefficient $\bar{\epsilon}$ has the value 0 for P inside surface S , the value 1 for P outside surface S , and the value 1 minus the solid angle of point P divided by 4π for P on surface S . $\bar{\epsilon}$ has the value of $1/2$ for P on the smooth part of S . G is the free-space Green's function for the Helmholtz equation in which the time harmonic $e^{-i\omega t}$, is omitted, and can be written as

$$G(P, Q) = \frac{e^{ikr}}{r(P, Q)}, \quad (2)$$

where k is the wave number and r is the distance between point P and point Q . It is well-known that such an integral equation becomes singular at resonant wave numbers for the related interior problem and the solution is no longer unique. The remedy to overcome this problem of nonuniqueness, proposed by Burton and Miller,⁷ is to introduce a second integral equation by differentiating Eq. (1) in the normal direction at P :

$$\bar{\epsilon} \frac{\partial \Phi(P)}{\partial n_P} = \frac{\partial}{\partial n_P} \int_S \Phi(Q) \frac{\partial G}{\partial n_Q} dS_Q - \int_S \frac{\partial G}{\partial n_P} \frac{\partial \Phi}{\partial n_Q}(Q) dS_Q + \frac{\partial \Phi_{in}(P)}{\partial n_P}. \quad (3)$$

From now on, the value of $\bar{\epsilon}$ is taken as $1/2$ when P is on surface S , because we evaluate the velocity potentials at points on the smooth surface only. By combining the surface Helmholtz integral equation and its normal derivative equation, Burton and Miller show that a unique solution can be obtained for all frequencies:

$$\begin{aligned} & \frac{1}{2} \left[\Phi(P) + \alpha \frac{\partial \Phi(P)}{\partial n_P} \right] \\ &= \int_S \left[\Phi(Q) \frac{\partial G}{\partial n_Q} - G \frac{\partial \Phi}{\partial n_Q}(Q) \right] dS_Q \\ &+ \alpha \frac{\partial}{\partial n_P} \int_S \Phi(Q) \frac{\partial G}{\partial n_Q} dS_Q - \alpha \int_S \frac{\partial G}{\partial n_P} \frac{\partial \Phi}{\partial n_Q} dS_Q \\ &+ \Phi_{in}(P) + \alpha \frac{\partial \Phi_{in}(P)}{\partial n_P}, \end{aligned} \quad (4)$$

where α is a coupling constant that should be chosen such that

$$\begin{aligned} \text{Im}(\alpha) &\neq 0 \quad \text{for } k\text{-real or imaginary} \\ &= 0 \quad \text{for } k \text{ complex.} \end{aligned} \quad (5)$$

B. Potential theory

The integral representation corresponding to an interior Laplace problem with suitable boundary conditions on surface S is expressed as

$$\begin{aligned} \epsilon \Phi_0(P) &= - \int_S \left[\Phi_0(Q) \frac{\partial G_0(P, Q)}{\partial n_Q} \right. \\ &\quad \left. - G_0(P, Q) \frac{\partial \Phi_0}{\partial n_Q}(Q) \right] dS_Q, \end{aligned} \quad (6)$$

where $\Phi_0(Q)$ and $\partial \Phi_0(Q)/\partial n_Q$ are the velocity potential and the normal velocity on the boundary S , respectively. Here, ϵ satisfies the relation

$$\bar{\epsilon} + \epsilon = 1, \quad (7)$$

and $G_0(P, Q)$ is the free-space Green's function for the Laplace equation.

$$G_0(P, Q) = 1/4\pi r. \quad (8)$$

If $\Phi_0(Q)$ is a constant, then we have

$$\epsilon = - \int_S \frac{\partial G_0(P, Q)}{\partial n_Q} dS_Q. \quad (9)$$

Equation (9) is interpreted as Gauss' flux theorem in the potential theory and this known analytical form can be used to reduce some normal derivative type singularity for Helmholtz integral equations. Since the value of ϵ is constant, the directional derivative of ϵ is always zero inside the surface

S . The right-hand side integral of Eq. (9) is a double layer distribution in the potential theory. According to the double layer potential theory, the normal derivative of a double layer potential is continuous across the boundary if the double distribution is continuous in surface S .^{19,20} Therefore, $\partial\epsilon/\partial n_P$ is zero when P is on the surface S , since it is zero when P is inside the surface.

Another important property, which is useful to regularize the $1/r$ type singularity in the integral equation, is the source distribution for the equipotential surface from the potential theory.^{19,20} The value of equipotential, Φ_e , inside the domain can be calculated by an integral of the corresponding source density $S(Q)$:

$$\Phi_e = - \int_S \frac{S(Q)}{4\pi r} dS_Q. \quad (10)$$

The source density $S(Q)$ can be solved by an intermediate numerical step during the computation of the Helmholtz integral equation.¹⁸ By multiplying Eq. (9) by $\Phi(P)$ and taking the normal derivative of the product with respect to $\partial/\partial n_P$ and using the property $\partial\epsilon/\partial n_P=0$ when P is on the surface S , we obtain

$$\frac{\partial}{\partial n_P} \int_S \phi(P) \frac{\partial G_0(P, Q)}{\partial n_Q} dS_Q = -\epsilon \frac{\partial \phi(P)}{\partial n_P}. \quad (11)$$

Equation (11) is a very useful formula to reduce the hypersingular kernel in the normal derivative Helmholtz integral equation into a regular form.

II. REGULARIZATION OF SINGULAR KERNELS

The most difficult part to evaluate in Eq. (3) is the first term in the right-hand side, in which the kernel is strongly singular ($1/r^2$ in 3-D) as point Q approaches the point P . One should not move the differential operator $\partial/\partial n_P$ into the integral directly since a hypersingular kernel is nonintegrable. A special treatment, proposed by Chien *et al.*,¹² is needed to take the differential operator into the integral. Therefore, this term can be transformed into an integrable form by subtracting Eq. (11), and then applying the differential operator to the new integrand, since the new integrand is weakly singular at most:

$$\begin{aligned} & \frac{\partial}{\partial n_P} \int_S \Phi(Q) \frac{\partial G(P, Q)}{\partial n_Q} dS_Q \\ &= \frac{\partial}{\partial n_P} \int_S \left[\Phi(Q) \frac{\partial G}{\partial n_Q} - \Phi(P) \frac{\partial G_0}{\partial n_Q} \right] dS_Q - \epsilon \frac{\partial \Phi}{\partial n_P} \\ &= \int_S \left[\Phi(Q) \frac{\partial^2 G(P, Q)}{\partial n_P \partial n_Q} - \frac{\partial^2 G_0(P, Q)}{\partial n_P \partial n_Q} \Phi(P) \right. \\ & \quad \left. - \frac{\partial G_0}{\partial n_Q} \frac{\partial \Phi}{\partial n_P} \right] dS_Q - \epsilon \frac{\partial \Phi}{\partial n_P}. \end{aligned} \quad (12)$$

By using Eq. (9), the last term of Eq. (12) cancels the integration of the third term in the bracket of Eq. (12). Therefore,

$$\begin{aligned} & \frac{\partial}{\partial n_P} \int_S \Phi(Q) \frac{\partial G(P, Q)}{\partial n_Q} dS_Q \\ &= \int_S \left[\Phi(Q) \frac{\partial^2 G(P, Q)}{\partial n_P \partial n_Q} - \frac{\partial^2 G_0(P, Q)}{\partial n_P \partial n_Q} \Phi(P) \right] dS_Q. \end{aligned} \quad (13)$$

Substituting Eq. (13) into Eq. (4), we have

$$\begin{aligned} & \frac{1}{2} \left(\Phi(P) + \alpha \frac{\partial \Phi(P)}{\partial n_P} \right) - \left(\Phi_{in} + \alpha \frac{\partial \Phi_{in}}{\partial n_P} \right) \\ &= \int_S \left[\Phi(Q) \frac{\partial G}{\partial n_Q} - G \frac{\partial \Phi}{\partial n_Q}(Q) \right] dS_Q \\ & \quad + \alpha \int_S \left[\Phi(Q) \frac{\partial^2 G}{\partial n_P \partial n_Q} - \Phi(P) \frac{\partial^2 G_0}{\partial n_P \partial n_Q} \right. \\ & \quad \left. - \frac{\partial G}{\partial n_P} \frac{\partial \Phi}{\partial n_Q}(Q) \right] dS_Q. \end{aligned} \quad (14)$$

Since the integrand of Eq. (14) contains singularities $1/r$, $1/r^2$, and $1/r^3$ when point Q approaches point P , the direct computation of those singular functions is not available and some intermediate steps must be executed for numerical purpose. Regarding to the $1/r$ and $1/r^2$ type of singularities, they have been numerically implemented with much success by several researchers in the literature, and a different regularization step is discussed by Hwang¹⁸ with the assumption of twice continuous differentiable regular elements in the sense of Kellogg.¹⁹ This assumption is still used in the following discussion.

Although the order of singular kernel in the left-hand side integral of Eq. (13) can be reduced by the subtraction of two hypersingular functions, the kernel in the right-hand side of Eq. (13) is still strongly singular. A further regularization is introduced to reduce the strongly singular kernel into a bounded kernel by the following steps:

$$\begin{aligned} & \int_S \left[\Phi(Q) \frac{\partial^2 G}{\partial n_P \partial n_Q} - \frac{\partial^2 G_0}{\partial n_P \partial n_Q} \Phi(P) \right] dS_Q \\ &= \int_S \left[\frac{\partial^2 G}{\partial n_P \partial n_Q} - \frac{\partial^2 G_0}{\partial n_P \partial n_Q} \right] \Phi(Q) dS_Q \\ & \quad + \int_S \frac{\partial^2 G_0}{\partial n_P \partial n_Q} [\Phi(Q) - \Phi(P)] dS_Q. \end{aligned} \quad (15)$$

The second integral in the right-hand side of Eq. (15) is corresponding to an associate interior problem of the Laplace equation, $\nabla^2 \Phi_0 = 0$, with the boundary condition $\Phi_0 = \Phi$ on surface S . In order to calculate this integral, it is better to convert this integral into a different form, by using the potential theory and Eq. (13). However, the Laplace equation is just a special case of the Helmholtz equations, wave number $k = 0$. Therefore, the simplest way to accomplish this transformation is to replace the function $G(P, Q)$ by function $G_0(P, Q)$ in Eq. (13), and the modified expression is shown in Eq. (16):

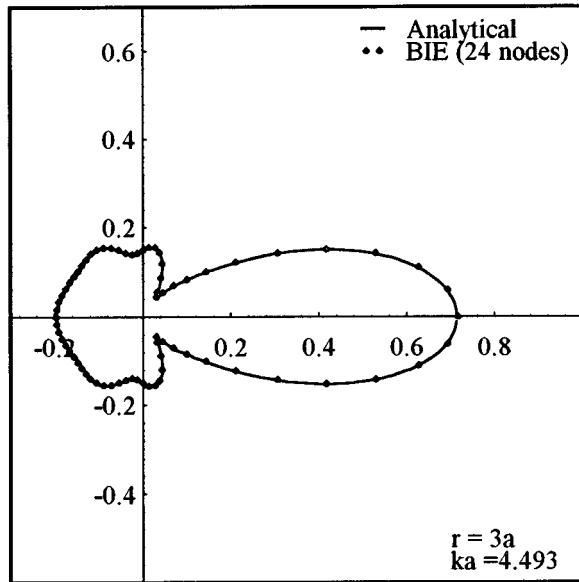


FIG. 1. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka=4.4934$ and $r=3a$.

$$\begin{aligned} \frac{\partial}{\partial n_P} \int_S \Phi_0(Q) \frac{\partial G_0}{\partial n_Q} dS_Q \\ = \int_S [\Phi_0(Q) - \Phi_0(P)] \frac{\partial^2 G_0}{\partial n_P \partial n_Q} dS_Q. \end{aligned} \quad (16)$$

Equation (16) establishes the relationship between the last integral in Eq. (15) and the associate Laplace problem. The left-hand-side integral in the Eq. (16) is a familiar term in the Green's identity. By differentiating Eq. (6) with respect to the operator $\partial/\partial n_P$, we have

$$\begin{aligned} \epsilon \frac{\partial}{\partial n_P} \Phi_0(P) = \int_S \frac{\partial G_0}{\partial n_P} \frac{\partial \Phi_0(Q)}{\partial n_Q} dS_Q \\ - \frac{\partial}{\partial n_P} \int_S \Phi_0(Q) \frac{\partial G_0}{\partial n_Q} dS_Q. \end{aligned} \quad (17)$$

Substituting the results in (16) and (17) into the last integral of Eq. (15), and utilizing the relation $\Phi_0 = \Phi$ on S and flux theorem (9), one obtains

$$\begin{aligned} \int_S [\Phi(Q) - \Phi(P)] \frac{\partial^2 G_0}{\partial n_P \partial n_Q} dS_Q \\ = \int_S \frac{\partial G_0}{\partial n_P} \frac{\partial \Phi_0(Q)}{\partial n_Q} dS_Q - \epsilon \frac{\partial \Phi_0(P)}{\partial n_P} \\ = \int_S \left[\frac{\partial G_0}{\partial n_P} \frac{\partial \Phi_0}{\partial n_Q} - \frac{\partial G_0}{\partial n_Q} \frac{\partial \Phi_0}{\partial n_P} \right] dS_Q - 2\epsilon \frac{\partial \Phi_0(P)}{\partial n_P}. \end{aligned} \quad (18)$$

It is already shown by Hwang¹⁸ that the last integrand of Eq. (18) is bounded but discontinuous, if the derivative of $\partial \Phi_0(P)/\partial n_P$ is continuous in the neighborhood of the point P . Again, $\partial \Phi_0(P)/\partial n_P$ can be solved in terms of $\Phi_0(P)$ through Eq. (6). For the convenience of computation, the left-hand-side term of Eq. (6) is replaced by Eq. (9) and

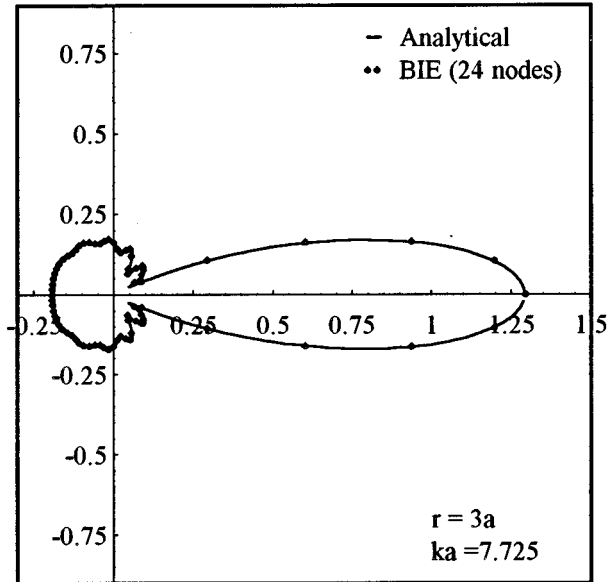


FIG. 2. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka=7.725$ and $r=3a$.

combined with the second part of the integral in Eq. (6). Then, by multiplying Eq. (10) with a factor $-(\partial \Phi_0/\partial n_P)/S(P)$, and adding it to the first part of the integral in Eq. (6), the final form of Eq. (6) becomes

$$\begin{aligned} \int_S G_0 \left[\frac{\partial \Phi_0}{\partial n_Q} - \frac{S(Q)}{S(P)} \frac{\partial \Phi_0}{\partial n_P} \right] dS_Q - \frac{\Phi_e}{S(P)} \frac{\partial \Phi_0}{\partial n_P} \\ = \int_S [\Phi_0(Q) - \Phi_0(P)] \frac{\partial G_0(P, Q)}{\partial n_Q} dS_Q. \end{aligned} \quad (19)$$

The integrands on both sides of Eq. (19) are bounded but discontinuous,¹⁸ therefore, the computation of integration is very easy to implement. After Eq. (19) is solved, $\partial \Phi_0(P)/\partial n_P$ is in terms of $\Phi(P)$ numerically, since $\Phi_0(P) = \Phi(P)$ on surface S .

As point Q approaches point P , the kernel of the first integral on the right-hand side of Eq. (15) can be expanded by Taylor series.

$$\frac{\partial^2 (G - G_0)}{\partial n_P \partial n_Q} = \frac{k^2}{8\pi r} + \frac{ik^3}{12\pi} + O(r). \quad (20)$$

Since the above kernel is weakly singular, its integral can be further regularized by subtracting Eq. (10) [multiply a factor $k^2 \Phi(P)/2S(P)$] from Eq. (20). The result is expressed as

$$\begin{aligned} \int_S \frac{\partial^2 (G - G_0)}{\partial n_P \partial n_Q} \Phi(Q) dS_Q \\ = \int_S \left[\Phi(Q) \frac{\partial^2 (G - G_0)}{\partial n_P \partial n_Q} - \frac{k^2 S(Q)}{2S(P)} G_0 \Phi(P) \right] dS_Q \\ - \frac{k^2 \Phi_e}{2S(P)} \Phi(P). \end{aligned} \quad (21)$$

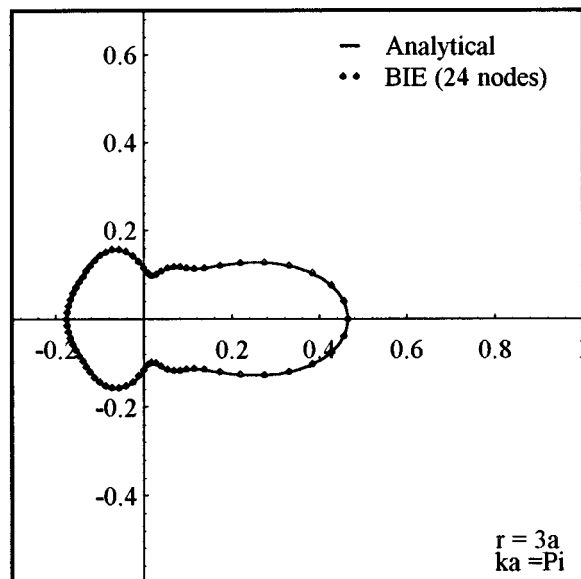


FIG. 3. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka = \pi$ and $r = 3a$.

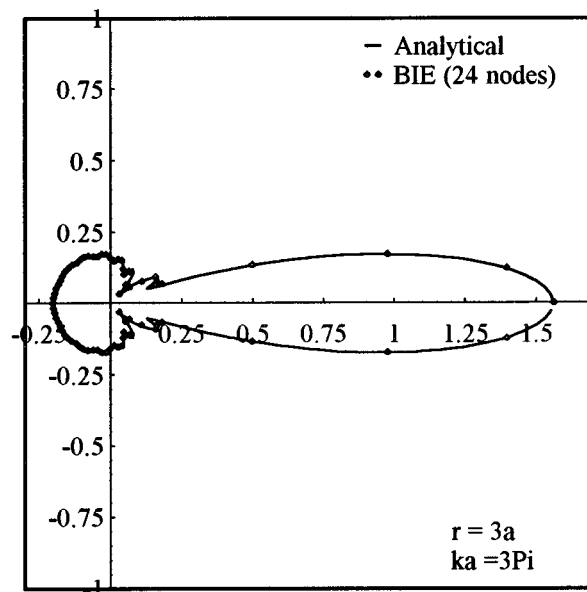


FIG. 4. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka = 3\pi$ and $r = 3a$.

As point Q approaches point P , the integrand in the right-hand side of Eq. (21) approaches $ik^3\Phi(P)/3$ in the average sense, if the first-order derivatives of $\Phi(P)$ and $S(P)$ are continuous in the neighborhood of point P . Substituting Eqs. (15), (18), and (21) into Eq. (14), we obtain

$$\begin{aligned} & \left[1 + \frac{\alpha k^2 \Phi_e}{2S(P)} \right] \Phi(P) - \frac{\Phi_e}{S(P)} \frac{\partial \Phi(P)}{\partial n_P} \\ &= \left(\Phi_{in} + \alpha \frac{\partial \Phi_{in}}{\partial n_P} \right) + \int_S \left\{ \left[\Phi(Q) \frac{\partial G}{\partial n_Q} - \frac{\partial G_0}{\partial n_Q} \Phi(P) \right] \right. \\ & \quad + \left[G_0 \frac{S(Q)}{S(P)} \frac{\partial \Phi}{\partial n_P} - G \frac{\partial \Phi}{\partial n_Q}(Q) \right] \left. \right\} dS_Q + \alpha \int_S \left\{ \left[\Phi(Q) \frac{\partial^2 (G - G_0)}{\partial n_P \partial n_Q} - \frac{k^2 S(Q)}{2S(P)} G_0 \Phi(P) \right] \right. \\ & \quad + \left. \left[\frac{\partial G_0}{\partial n_Q} \frac{\partial \Phi}{\partial n_P} - \frac{\partial G}{\partial n_P} \frac{\partial \Phi}{\partial n_Q} \right] \right\} dS_Q + \alpha \int_S \left[\frac{\partial G_0}{\partial n_P} \frac{\partial \Phi_0}{\partial n_Q} - \frac{\partial G_0}{\partial n_Q} \frac{\partial \Phi_0}{\partial n_P} \right] dS_Q + \alpha \left(\frac{1}{2} \frac{\partial \Phi}{\partial n_P} - \frac{\partial \Phi_0}{\partial n_P} \right). \end{aligned} \quad (22)$$

In Eq. (22), all the kernels of surface integrals are bounded, therefore, the standard Gaussian quadrature formula can be directly applied without any modification.

III. NUMERICAL EXAMPLES

The numerical example is the scattering problem of a rigid sphere due to the plane incident wave. The incoming unit plane wave travels to the right along the x axis and is described as $\Phi_0 = \exp(ikx)$. For the scattering wave from a sphere of radius a , the scattered velocity potential at a distance r from the center of the sphere and an angle θ from the x axis is given by²¹

$$\Phi_S(r, \theta) = \sum_{n=1}^{\infty} -\frac{i^n (2n+1) j'_n(ka)}{h'_n(ka)} P_n(\cos \theta) h_n(kr), \quad (23)$$

where P_n is the Legendre function of the first kind, h_n the spherical Hankel function of the first kind, and j_n the spherical Bessel function of the first kind.

The magnitudes of scattered velocity potential at a distance $r = 3a$ are plotted versus the polar angle θ for nondimensional wave numbers, and compared with the analytical solutions in Figs. 1–5. Due to the axisymmetry, the velocity potentials are solved along a semicircle, $\theta = 0$. In all the cases, the 24-point Gauss–Legendre formula is used, and the 24 Gaussian–Legendre nodes are the exact computing points. The coupling factor α for the composite integral equation is set to be i/k for convenience. Figures 1 and 2 show the results for $ka = 4.4934$ and 7.725 , respectively, which are the fictitious frequencies of the normal derivative boundary integral equations. Figures 3 and 4 show the results for $ka = \pi$ and 3π , respectively, which are the fictitious frequencies of both the conventional and the normal derivative boundary integral equations. Figure 5 shows the result at

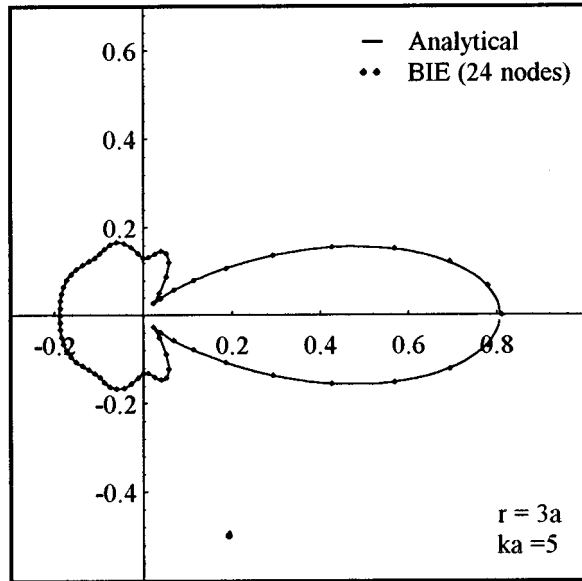


FIG. 5. Plot of $|\phi_s|$ versus polar angle for a rigid sphere when $ka=5$ and $r=3a$.

$ka=5$, which is not a fictitious frequency for both boundary integral equations. The agreement between the analytical solutions and the results of the proposed numerical method is excellent in all the cases.

An oscillating sphere is considered next. The analytical solution of velocity potential for an oscillating sphere of radius a with a radial velocity $\cos \theta$ is given by²²

$$\Phi(P) = \left(\frac{a}{r}\right)^2 \cos \theta \frac{a(1-ikr)(k^2a^2 - 2 - 2ika)}{k^4a^4 + 4} \times e^{ik(r-a)}. \quad (24)$$

Due to the axisymmetry, the velocity potentials are still solved along a semicircle, $\theta=0$. The 24-point Gauss-Legendre formula is used along the semicircle, and the 24 nodes are the exact collocation points. In Fig. 6, the magnitude of radiating velocity potential at the collocation points on the sphere surface is plotted versus the polar angle θ for nondimensional wave numbers, $ka=\pi$, and compared with

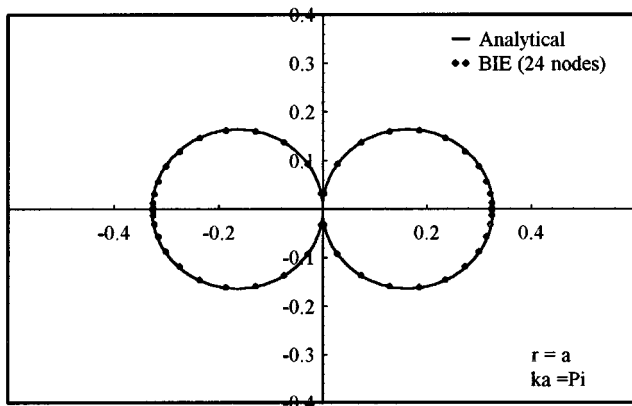


FIG. 6. Absolute value of radiating velocity potential $|\phi|$ versus polar angle for a rigid sphere when $ka=\pi$ and $r=a$.

the analytical solutions. Again, no significant error comes up in Fig. 6 as shown.

IV. CONCLUSIONS

It has been demonstrated that the present method to regularize the hypersingular integral in the Helmholtz integral equation is very accurate and efficient. All the computations of singular integral have been removed and accomplished with the standard Gaussian quadrature procedure without any special quadrature scheme or excessive quadrature points near singular points.

There is another important advantage in the present method. Since all the computations are executed at the Gauss-Legendre nodes, they do not locate at boundaries of regular surface elements. Therefore, no computation is necessary at those points on the nonsmooth edges or corners, in which the normal derivative form of the conventional Helmholtz equation is undefined. The computation at the nonsmooth surface is tedious for conventional boundary element methods usually.

ACKNOWLEDGMENTS

The results in this paper were partially supported by the National Science Council, R.O.C., under Grant No. NSC 86-2611-E-002-031.

- ¹H. A. Schenck, "Improved integral formulation for acoustic radiation problems," *J. Acoust. Soc. Am.* **44**, 41–58 (1968).
- ²A. F. Seybert and T. K. Rengarajan, "The use of CHIEF to obtain unique solution for acoustic-radiation using boundary integral equations," *J. Acoust. Soc. Am.* **81**, 1299–1306 (1987).
- ³J. Zhou and H-Z. Wang, "Calculating acoustic radiation from closed bodies using BIE Method," Proceedings of the 2nd Marine and Weapon Underwater Noise Meeting of China, Yangzhou, 1987.
- ⁴T. W. Wu and A. F. Seybert, "Acoustic radiation and scattering, in *Advanced Boundary Element Methods in Acoustics*, edited by R. D. Ciskowski and C. A. Brebia (Computational Mechanics, Southampton, U.K., 1991), Chap. 3.
- ⁵T. W. Wu and A. F. Seybert, "A weighted residual formulation for the CHIEF methods in acoustics," *J. Acoust. Soc. Am.* **90**, 1608–1614 (1991).
- ⁶D. J. Segalman and D. W. Lobitz, "A method to overcome computational difficulties in the exterior acoustics problem," *J. Acoust. Soc. Am.* **91**, 1855–1861 (1992).
- ⁷A. J. Burton and G. F. Miller, "The application of integral equation methods to the solution of some exterior boundary-value problems," *Proc. R. Soc. London, Ser. A* **323**, 201–210 (1971).
- ⁸S. Amini and D. T. Wilton, "An investigation of boundary element methods for the exterior acoustic problem," *Comput. Methods Appl. Mech. Eng.* **54**, 49–65 (1986).
- ⁹W. L. Meyer, W. A. Bell, B. T. Zinn, and M. P. Stallybrass, "Boundary integral solutions of three dimensional acoustic radiation problems," *J. Sound Vib.* **59**, 245–262 (1978).
- ¹⁰I. C. Mathews, "Numerical techniques for three-dimensional steady-state fluid-structure interaction," *J. Acoust. Soc. Am.* **79**, 1317–1325 (1986).
- ¹¹T. Terai, "On calculation of sound fields around three dimensional objects by integral equation methods," *J. Sound Vib.* **69**, 71–100 (1980).
- ¹²C. C. Chien, J. Rajiyah, and S. N. Atluri, "An effective method for solving the hypersingular integral equations in 3-D acoustics," *J. Acoust. Soc. Am.* **88**, 918–937 (1990).
- ¹³G. Krishnasamy, L. W. Schmerr, T. J. Rudolph, and F. J. Rizzo, "Hypersingular boundary integral equations: some applications in acoustic and elastic wave scattering," *J. Appl. Mech.* **57**, 404–414 (1990).
- ¹⁴Yijun-Liu and F. J. Rizzo, "A weakly singular form of the hypersingular boundary integral equation applied to 3-D acoustic wave problems," *Comput. Methods Appl. Mech. Eng.* **96**, 271–287 (1992).

- ¹⁵T. W. Wu, A. F. Seybert, and G. C. Wan, "On the numerical implementation of a Cauchy principal value integral to insure a unique solution for acoustic radiation and scattering," *J. Acoust. Soc. Am.* **90**, 554–560 (1991).
- ¹⁶T. W. Wu and G. C. Wan, "Numerical modeling of acoustic radiation and scattering from thin bodies using a Cauchy principal integral equation," *J. Acoust. Soc. Am.* **92**, 2900–2906 (1992).
- ¹⁷A. W. Maue, "Zur formulierung eines allgemeinen beugungsproblems durch eine integralgleichung," *Z. Phys.* **126**, 601–618 (1949).
- ¹⁸W. S. Hwang, "A boundary integral method for acoustic radiation and scattering," *J. Acoust. Soc. Am.* **101**, 3330–3335 (1997).
- ¹⁹O. D. Kellogg, *Foundations of Potential Theory* (Dover, New York, 1954).
- ²⁰M. A. Jaswon and G. T. Symm, *Integral Equation Methods in Potential Theory and Elastostatics* (Academic, London, 1977).
- ²¹R. Hickling and N. M. Wang, "Scattering of sound by a rigid movable sphere," *J. Acoust. Soc. Am.* **39**, 276–279 (1966).
- ²²A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981).

Sound propagation from a dipole source near an impedance plane

Kai Ming Li, Shahram Taherzadeh, and Keith Attenborough

Engineering Mechanics Discipline, Faculty of Technology, The Open University, Milton Keynes MK7 6AA, United Kingdom

(Received 5 December 1995; accepted for publication 12 February 1997)

A closed-form analytic solution has been derived for an arbitrarily oriented dipole placed above an impedance plane. Asymptotic approximations for the total sound field can be written in a form similar to the classical formula for a monopole. Analytic approximations derived for horizontal and vertical dipoles give predictions that agree well with those obtained from a program of the fast field type. The asymptotic results have been confirmed also by laboratory measurements. Compared with numerical methods, the analytic approximations offer the combined advantages of easier physical interpretation and a much reduced computational time. It is found that the variation of the sound-pressure level from a vertical dipole above an impedance plane differs significantly from that due to a monopole above this plane, particularly as a result of differences in the ground wave component. However, the excess attenuation due to a horizontal dipole is found to be rather similar to that for a monopole except for source and receiver heights that are comparable with range. © 1997 Acoustical Society of America. [S0001-4966(97)02706-9]

PACS numbers: 43.28.Fp, 43.55.Ev, 43.58.Bh [LCS]

INTRODUCTION

There has been sustained interest in the theory of sound propagation above an impedance ground for many years. The sound field due to a monopole source in a homogenous medium above an impedance ground is cylindrically symmetric so that there is no azimuthal variation. The theory for this situation is well-known¹ and has proved to be very useful for many applications of outdoor sound propagation. However, many noise sources are directional and do not behave as simple monopoles particularly at close range. Source descriptions in terms of multipoles may be more useful. Although multipole sources radiate a much weaker sound field than the corresponding monopole source, there are many situations when the multipole source strength is very large and there is no significant contribution from monopole radiation. For instance, sound radiated by an oscillating sphere or an unenclosed loudspeaker can be represented by a dipole.² Rolling noise from a train has been characterized as that from an incoherent line of dipoles.³ Lighthill's acoustic analogy⁴ establishes that the sources of a quadrupole nature may be used to model jet noise. The sound fields due to these multipole sources are no longer cylindrically symmetric but depend on their orientations. Consequently predictions of the sound field due to multipole noise sources above an impedance ground should be important in noise assessment and control. They may prove useful also in passive source identification and in near-field holography.⁵

Despite the existence of substantial literature on the theory of the dipole radiation above a boundary plane in the electromagnetic scattering theory,^{6,7} little attention has been devoted to equivalent situations in atmospheric and underwater acoustics. Generalov⁸ has derived an asymptotic approximation for the sound field due to both a dipole and a quadrupole near a locally reacting boundary. Hu and Bolton⁹ have computed the corresponding sound field numerically by

means of a two-dimensional Hankel transform and have carried out laboratory experiments on propagation from an un baffled loudspeaker over porous surfaces. Their numerical results do not agree well with Generalov's asymptotic solution. Moreover, Hu and Bolton found it necessary to include a substantial monopole component in their predictions to obtain agreement between their calculations and experiments. Nevertheless, their results indicate that the presence of a dipole component has a significant effect on the total field including diffraction and ground reflection.

The principal objective of this paper is to derive an asymptotic expression for the sound field due to an arbitrarily oriented multipole source above an impedance plane. Other factors which affect the propagation of sound outdoors, such as turbulence, wind and temperature gradients, and air-absorption, will not be considered here. In Sec. I, we offer an analysis from first principles. A multipole source of an arbitrary orientation is incorporated in the governing Helmholtz equation. We limit our attention here to the solution for dipole sources, however, our procedure can be extended easily to multipoles of a higher order. To derive asymptotic approximations, we use the method of steepest descents.^{10,11} Calculations have been carried out also by a numerical method detailed elsewhere¹² and, in Sec. II, are used to verify the accuracy of the asymptotic expressions. In Sec. III, we describe laboratory experiments with a dipole source and compare theoretical predictions with data, before making a few concluding remarks.

I. THEORY

A. Formulation

Consider a multipole source with its center at a height of z_s above a semi-infinite impedance ground, with the source–receiver configuration shown in Fig. 1. The total sound field

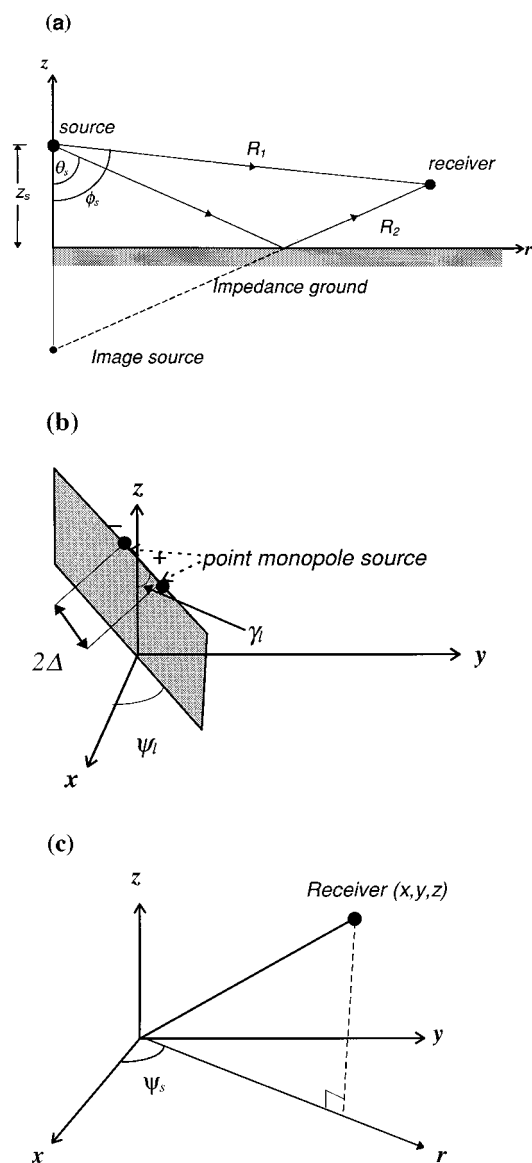


FIG. 1. Schematic diagram for the source-receiver geometry and the orientation of the dipole source.

$p(x,y,z)$ satisfies the inhomogeneous Helmholtz equation and an impedance boundary condition at the ground surface, $z=0$. The governing inhomogeneous Helmholtz equation is

$$\nabla^2 p + k^2 p = \Gamma_j(\mathbf{r}_s), \quad j=0,1,2,\dots, \quad (1)$$

where Γ_j Pa m⁻² represents the multipole source term at the position $\mathbf{r}_s \equiv (0,0,z_s)$ and the subscript j denotes the order of the multipole. A monopole is given by $j=0$, a dipole by $j=1$, and a quadrupole by $j=2$, etc. Once the source term Γ_j has been specified, we can solve Eq. (1) to give the sound field due to a multipole. As the source and receiver are close to an impedance ground, the sound pressure p must also satisfy the boundary condition on the plane $z=0$, i.e.,

$$\frac{\partial}{\partial z} p(x,y,0) + ik\beta p(x,y,0) = 0, \quad (2)$$

where β is the normalized admittance of the ground.

The source term for a monopole source of strength S_0 Pa m is given by

$$\Gamma_0(\mathbf{r}_s) = -S_0 \delta(x) \delta(y) \delta(z - z_s).$$

The source term Γ_1 for an arbitrarily oriented dipole, with 2Δ m between the two component monopoles, is given by

$$\Gamma_1(\mathbf{r}_s) = \Gamma_0(\mathbf{r}_s + \mathbf{l}\Delta) - \Gamma_0(\mathbf{r}_s - \mathbf{l}\Delta) \quad \text{as } \Delta \rightarrow 0,$$

or

$$\Gamma_1(\mathbf{r}_s) = S_0 [\delta(x + l_x \Delta) \delta(y + l_y \Delta) \delta(z - z_s + l_z \Delta) - \delta(x - l_x \Delta) \delta(y - l_y \Delta) \delta(z - z_s - l_z \Delta)],$$

where $\mathbf{l} \equiv (l_x, l_y, l_z)$ are the direction cosines of the dipole axis which characterize the orientation. The expression for the dipole can be written as

$$\Gamma_1(\mathbf{r}_s) = S_1 \left[l_x \delta(y) \delta(z - z_s) \frac{\partial}{\partial x} \delta(x) + l_y \delta(x) \delta(z - z_s) \frac{\partial}{\partial y} \delta(y) + l_z \delta(x) \delta(y) \frac{\partial}{\partial z} \delta(z - z_s) \right], \quad (3)$$

where $S_1 = 2S_0\Delta$ Pa m², is the dipole source strength. It is found more convenient to use spherical polar coordinates in our subsequent analyses and they are given by

$$l_x = \sin \gamma_l \cos \psi_l, \quad (4a)$$

$$l_y = \sin \gamma_l \sin \psi_l, \quad (4b)$$

and

$$l_z = \cos \gamma_l, \quad (4c)$$

where γ_l is the polar angle measured from the negative z axis and ψ_l is the azimuthal angle measured from the positive x axis [Fig. 1(b) and (c)]. In addition, ψ_s is the azimuthal angle (angle measured in the x - y plane) of the line joining the source and receiver.

Expressions for Γ_j corresponding to higher-order multipole sources can be derived in a similar manner. Pierce¹⁴ gives a detailed description of methods for calculating the sound field due to a multiple source in an unbounded medium. Here, we limit our analysis to dipole sources and proceed to calculate sound propagation from an arbitrarily oriented dipole above an impedance plane.

B. Prediction of the sound field

For a dipole, the source term in Eq. (1) is replaced by that given by Eq. (3). The resulting governing equation can be solved by the method of Fourier transformation. Consider the Fourier transform pair, $p(x,y,z)$ and $\hat{p}(k_x, k_y, z)$ such that

$$\hat{p}(k_x, k_y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x,y,z) e^{-i(k_x x + k_y y)} dx dy \quad (5)$$

and

$$p(x,y,z) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{p}(k_x, k_y, z) e^{i(k_x x + k_y y)} dk_x dk_y. \quad (6)$$

Using Eq. (3) for the source term and applying Eq. (5) into Eqs. (1) and (2), the transformed governing equation becomes

$$\frac{d^2 \hat{p}}{dz^2} + k_z^2 \hat{p} = -S_1 \left[i(k_x l_x + k_y l_y) \delta(z - z_s) - l_z \frac{\partial}{\partial z} \delta(z - z_s) \right] \quad \text{for } z > 0, \quad (6)$$

and

$$\frac{\partial \hat{p}}{\partial z} + ik \beta \hat{p} = 0 \quad \text{on } z = 0. \quad (7)$$

We require also that the solution satisfies the Sommerfeld radiation condition at infinity

$$\hat{p} = \hat{\Gamma}_1^- e^{ik_z |z - z_s|} + \hat{\Gamma}_1^+ V e^{ik_z (z_s + z)}, \quad (8)$$

where

$$\hat{\Gamma}_1^- = \frac{S_1}{2k_z} [k_x l_x + k_y l_y + k_z l_z], \quad (9a)$$

$$\hat{\Gamma}_1^+ = \frac{S_1}{2k_z} [k_x l_x + k_y l_y + \text{sgn}(z - z_s) k_z l_z], \quad (9b)$$

$$V = \frac{k_z - k\beta}{k_z + k\beta}, \quad (10)$$

and

$$k_z = +\sqrt{k^2 - k_x^2 - k_y^2}. \quad (11)$$

The positive root of k_z is chosen in Eq. (11) in order to ensure a finite and bounded solution for the total sound field. The form of the transformed pressure given in Eq. (8) is similar to that of the transformed pressure due to a monopole.¹³ The variables $\hat{\Gamma}_1^-$ and $\hat{\Gamma}_1^+$ are parameters that characterize the nature of the source, which is a dipole in the present situation. It is not difficult to show that the corresponding parameters for the monopole are given by $\hat{\Gamma}_0^- = \hat{\Gamma}_0^+ = S_0/2ik_z$.

From Eqs. (6) and (8), the total sound field can be obtained by inverting the Fourier transform to yield

$$p(x, y, z) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\hat{\Gamma}_1^- e^{ik_z |z - z_s|} + \hat{\Gamma}_1^+ V e^{ik_z (z_s + z)}] e^{i(k_x x + k_y y)} dk_x dk_y. \quad (12)$$

The expression for the total sound field can be simplified considerably by using cylindrical polar coordinates such that (k_x, k_y) become (k_r, ξ) and (x, y) become (r, ψ_s) . Here, k_r and r represent the corresponding magnitudes while ξ and ψ_s represent the corresponding azimuthal angles. A close examination of Eqs. (8) and (12) reveals that the total sound field can be expressed as a sum of two parts: a horizontal-dipole component (when $\gamma_L = \pi/2$) and a vertical-dipole component (when $\gamma_L = 0$), respectively. Making use of polar coordinates and substituting Eqs. (8)–(11) into Eq. (12), we can rewrite the total sound field as

$$p = p_h + p_v, \quad (13)$$

where

$$p_h = \frac{S_1 \sin \gamma_L}{4\pi^2} \int_0^{\infty} \int_0^{2\pi} \frac{k_r^2 \cos(\xi - \psi_l)}{2k_z} [e^{ik_z |z_s - z|} + V e^{ik_z (z_s + z)}] e^{ik_r r \cos(\xi - \psi_s)} d\xi dk_r, \quad (14)$$

$$p_v = \frac{S_1 \cos \gamma_L}{4\pi^2} \int_0^{\infty} \int_0^{2\pi} \frac{k_r}{2} [\text{sgn}(z - z_s) e^{ik_z |z_s - z|} + V e^{ik_z (z_s + z)}] e^{ik_r r \cos(\xi - \psi_s)} d\xi dk_r, \quad (15)$$

with

$$k_z = +\sqrt{k^2 - k_r^2}. \quad (16)$$

The integral over the azimuthal angle for the horizontal-dipole and vertical-dipole components, Eqs. (14) and (15), respectively, can be evaluated immediately to give

$$p_h = \frac{-iS_1 \sin \gamma_L \cos(\psi_s - \psi_l)}{4\pi} \int_0^{\infty} \frac{k_r^2}{k_z} [e^{ik_z |z_s - z|} + V e^{ik_z (z_s + z)}] J_1(k_r r) dk_r, \quad (17)$$

$$p_v = \frac{S_1 \cos \gamma_L}{4\pi} \int_0^{\infty} k_r [\text{sgn}(z - z_s) e^{ik_z |z_s - z|} + V e^{ik_z (z_s + z)}] J_0(k_r r) dk_r. \quad (18)$$

In the evaluations of integrals over ξ , the following identities¹⁵ have been used:

$$J_n(z) = \frac{i^{-n}}{2\pi} \int_0^{2\pi} \cos(n\psi) e^{iz \cos \psi} d\psi \quad (19a)$$

and

$$\int_0^{2\pi} \sin(n\psi) e^{iz \cos \psi} d\psi = 0, \quad (19b)$$

where $J_n(z)$ is the n th-order Bessel function. The expressions given in Eqs. (17) and (18) agree with that derived by Hu and Bolton⁹ for a horizontal dipole and a vertical dipole, respectively. They are integral representations of the sound field and can be evaluated numerically by the Fourier transform method used in fast-field programs.¹² In this section, we seek an asymptotic evaluation of the inverse Fourier integrals starting with that for the horizontal-dipole component.

It is convenient to split the horizontal-dipole component further into three parts; p_1 , p_2 , and p_3 as follows:

$$p_1 = \frac{-iS_1 \sin \gamma_L \cos(\psi_s - \psi_l)}{4\pi} \int_0^{\infty} \frac{k_r^2}{k_z} e^{ik_z |z_s - z|} J_1(k_r r) dk_r, \quad (20a)$$

$$p_2 = \frac{-iS_1 \sin \gamma_L \cos(\psi_s - \psi_l)}{4\pi} \int_0^{\infty} \frac{k_r^2}{k_z} e^{ik_z (z_s + z)} J_1(k_r r) dk_r, \quad (20b)$$

and

$$p_3 = \frac{iS_1 k \beta \sin \gamma_l \cos(\psi_s - \psi_l)}{2\pi} \times \int_0^\infty \frac{k_r^2}{k_z(k_z + k\beta)} e^{ik_z(z_s + z)} J_1(k_r r) dk_r. \quad (20c)$$

The first term p_1 may be interpreted as the direct sound field due to the horizontal-dipole component, of an arbitrary dipole source, in an unbounded medium. The second term p_2 may be regarded as the sound field due to its corresponding image source above a perfectly reflecting plane. Finally, the third term p_3 is the correction term which takes into account the ground of finite impedance. In our subsequent analysis, we shall refer to these three components as the direct wave term, image wave term, and the boundary wave term, respectively.

The exact evaluations of the integrals for the direct and image wave terms, as given in Eqs. (20a) and (20b), pose considerable problems. It seems rather difficult, if not impossible to evaluate the integrals exactly. However, we note that these two terms correspond to the sound field due to a dipole source and its image in an unbounded medium. It is obvious that the well-known result in an unbounded medium¹⁴ can be used to give exact expressions for p_1 and p_2 as

$$p_1 = \frac{S_1 \sin \gamma_l \sin \phi_s \cos(\psi_s - \psi_l)}{4\pi} \left[\frac{1 - ikR_1}{R_1^2} \right] e^{ikR_1} \quad (21a)$$

and

$$p_2 = \frac{S_1 \sin \gamma_l \sin \theta_s \cos(\psi_s - \psi_l)}{4\pi} \left[\frac{-ikR_2}{R_2^2} \right] e^{ikR_2}, \quad (21b)$$

where R_1 and R_2 are the distances from the source and its image to the receiver, respectively. The angles ϕ_s and θ_s are the polar angles of the lines joining the source and the image source to receiver and are shown in Fig. 1(a). On the other hand, the exact solution for integral of the boundary wave term p_3 [see Eq. (20c)] cannot be found but it can be evaluated asymptotically by the double saddle-point method. Details of the asymptotic method are given elsewhere.^{1,6,11} Here we give only an outline of the analysis and its results.

In the present analysis, we have used the spherical polar coordinate system as it is used in the analogous analysis for a monopole source.^{1,6,10,11} Using the relation,¹⁵

$$J_n(k_r r) = \frac{1}{2} \{H_n^{(1)}(k_r r) - H_n^{(1)}(-k_r r)\}$$

and introducing a polar angle μ , we can rewrite Eq. (20c) as

$$p_3 = \frac{iS_1 k^2 \beta \sin \gamma_l \cos(\psi_s - \psi_l)}{4\pi} \times \int_{-\pi/2 + i\infty}^{\pi/2 - i\infty} \Xi_h(\mu) e^{ikR_2 \cos(\mu - \theta_s)} d\mu, \quad (22)$$

where

$$\Xi_h(\mu) = \frac{\sin^2 \mu}{\cos \mu + \beta} \{H_1^{(1)}(kr \sin \mu) e^{-ikr \sin \mu}\}, \quad (23)$$

$H_n^{(1)}(z)$ is the n th-order Hankel function and μ is the polar angle measured from the negative z axis to the line joining

the source to the receiver. We note that μ is not only limited to the real values of this angle ($0 \leq \mu \leq \pi$) but it is also allowed to take complex values in our analysis. The physical interpretation of this mathematical abstraction is that the real values of μ correspond to the expansion of the usual waves in all possible directions, whereas the complex angles correspond to inhomogeneous waves (see Sec. 1.2 of Ref. 10).

We use the integral representation of the Hankel function and its asymptotic expansion,¹⁶

$$H_n^{(1)}(z) e^{-iz} = \frac{(-1)^n}{\pi z} \frac{4^{1-n}}{(2n-1) \cdots 5 \times 3 \times 1} \times \int_0^\infty y^{2n} e^{-y^2/2} (4iz - y^2)^{n-1/2} dy \quad (24a)$$

and

$$H_n^{(1)}(z) e^{-iz} \approx (-i)^n \left(\frac{2}{i\pi z} \right)^{1/2} \left\{ 1 - \left(\frac{4n^2 - 1}{8iz} \right) + \cdots \right\}, \quad (24b)$$

where $n \geq 1$. A uniform asymptotic expression for Eq. (22) can then be obtained by the method of pole subtraction to give

$$p_3 = \frac{k^2 \beta}{4} S_1 \sin \gamma_l \cos(\psi_s - \psi_l) (1 - \beta^2)^{1/2} \times \operatorname{erfc}(-ix_0/\sqrt{2}) H_1^{(1)}\{kr(1 - \beta^2)^{1/2}\} e^{-ik(z+z_s)\beta}, \quad (25)$$

where

$$\frac{1}{2}x_0^2 = ikR_2\{1 + \beta \cos \theta_s - (1 - \beta^2)^{1/2} \sin \theta_s\}.$$

x_0 corresponds to the pole of the integrand in Eq. (22) which occurs in the complex μ plane at the point, $\mu_p = \pi/2 + \cos^{-1}\sqrt{1 - \beta^2}$. Hence

$$\cos \mu_p = -\beta \quad (26a)$$

and

$$\sin \mu_p = \sqrt{1 - \beta^2}. \quad (26b)$$

The asymptotic solution of the sound field due to the horizontal dipole component is obtained by combining Eqs. (21a), (21b), and (25) to give

$$p_h = \frac{S_1 \sin \gamma_l \cos(\psi_s - \psi_l)}{4\pi} \left\{ \sin \phi_s \left[\frac{1 - ikR_1}{R_1^2} \right] e^{ikR_1} + \sin \theta_s R_p \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2} + \sin \mu_p (1 - R_p) F(w) \times \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2} \right\}, \quad (27)$$

where

$$R_p = \frac{\cos \theta_s - \beta}{\cos \theta_s + \beta}, \quad (28a)$$

$$F(w) = 1 + i\pi^{1/2} w e^{-w^2} \operatorname{erfc}(-iw), \quad (28b)$$

$$w = (\frac{1}{2} ikR_2)^{1/2} (\cos \theta_s + \beta), \quad (28c)$$

and $\sin \mu_p$ is given by Eq. (26b).

In Eq. (27), the terms R_p , $F(w)$, and w may be identified as the plane-wave reflection coefficient, the boundary loss factor, and the numerical distance, respectively. For propagation at near grazing incidence over a relatively hard boundary, $\sin \theta_s \approx 1$ and $\sin \mu_p = \sqrt{1 - \beta^2} \approx 1$ which means that $\sin \theta_s \approx \sin \mu_p$. For the horizontal-dipole component, it is convenient to replace $\sin \mu_p$ by $\sin \theta_s$ because the sound field can then be expressed in the well-known classical form as

$$p_h = \frac{S_1 \sin \gamma_l \cos(\psi_s - \psi_l)}{4\pi} \left\{ \sin \phi_s \left[\frac{1 - ikR_1}{R_1^2} \right] e^{ikR_1} + Q \sin \theta_s \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2} \right\}, \quad (29)$$

where the complex reflection coefficient Q is given by

$$Q = R_p + (1 - R_p)F(w). \quad (30)$$

We emphasize that image wave and the boundary wave term, see Eqs. (21b) and (25), are combined to give the second term of the curly bracket in Eq. (29). For a monopole source, it is customary to refer to the first term of the complex reflection coefficient, Q , see Eq. (30), as the reflected wave term that takes into account the interaction of plane wave with an impedance ground and to the second term as the ground wave term. Note that the additional assumptions $\beta^2 \ll 1$, $r \gg \max(z_s, z)$ and $kR_2 \gg 1$ have been used to obtain Eqs. (27) and (29). These assumptions are also used in the derivation of the asymptotic solution for a monopole source.¹¹

Using the same method, we can derive the vertical-dipole component of the sound field due in a similar manner. Again, the vertical-dipole component of the sound field consists of three parts, the direct wave p_4 , the image wave p_5 , and the boundary wave term p_6 , similar to that of the horizontal-dipole component, cf. Eqs. (20a)–(20c). The direct and image waves may be identified and expressed as

$$p_4 = \frac{S_1 \cos \gamma_l \cos \phi_s}{4\pi} \left[\frac{1 - ikR_1}{R_1^2} \right] e^{ikR_1}, \quad (31a)$$

$$p_5 = \frac{S_1 \cos \gamma_l \cos \theta_s}{4\pi} \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2}. \quad (31b)$$

In addition, the boundary term p_6 can be written in an integral form as

$$p_6 = - \frac{S_1 k^2 \beta \cos \gamma_l}{4\pi} \times \int_{-\pi/2 + i\infty}^{\pi/2 - i\infty} \Xi_v(\mu) e^{ikR_2 \cos(\mu - \theta_s)} d\mu, \quad (32)$$

where

$$\Xi_v(\mu) = \left[1 - \frac{\beta}{\cos \mu + \beta} \right] \sin \mu \{ H_0^{(1)} \times (kr \sin \mu) e^{-ikr \sin \mu} \}. \quad (33)$$

It is tedious but straightforward to show that p_6 can be approximated asymptotically by

$$p_6 = \frac{iS_1 k \beta \cos \gamma_l}{4\pi} \{ 2 + (1 - R_p) i \pi^{1/2} w e^{-w^2} \times \operatorname{erfc}(-iw) \} \frac{e^{ikR_2}}{R_2}. \quad (34)$$

Hence the sound field due to a vertical dipole can be obtained by combining Eqs. (31a), (31b), and (34) to yield

$$p_v = \frac{S_1 \cos \gamma_l}{4\pi} \left\{ \cos \phi_s \left[\frac{1 - ikR_1}{R_1^2} \right] e^{ikR_1} + R_p \cos \theta_s \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2} + \cos \mu_p (1 - R_p) F(w) \times \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2} \right\}, \quad (35)$$

where the approximations of $\beta^2 \ll 1$, $r \gg \max(z_s, z)$ and $kR_2 \gg 1$ have been used in order to arrive at the above closed-form analytic solution. The asymptotic solution is composed of three terms, a direct wave, a reflected wave, and a ground wave term. Note again that the image wave and the boundary wave terms (p_5 and p_6 , respectively) have been combined to give the reflected wave and ground wave terms Eq. (35). It is tempting to replace μ_p with θ_s for the ground wave term as we suggested earlier for the horizontal-dipole component. This approximation does not give satisfactory agreement with the other numerical scheme described in Sec. II. The ground wave term is more significant for the vertical-dipole component.

We can now write an expression of the total sound field due to an arbitrarily oriented dipole in a compact form by summing the contributions due to the horizontal-dipole and vertical-dipole components [see Eqs. (27) and (35)]. The asymptotic solution can be written in a more compact form if we define the following unit vectors

$$\hat{\mathbf{R}}_1 \equiv (\sin \phi_s \cos \psi_s, \sin \phi_s \sin \psi_s, \cos \phi_s), \quad (36a)$$

$$\hat{\mathbf{R}}_2 \equiv (\sin \theta_s \cos \psi_s, \sin \theta_s \sin \psi_s, \cos \theta_s), \quad (36b)$$

and

$$\hat{\mathbf{R}}_s \equiv (\sin \mu_p \cos \psi_s, \sin \mu_p \sin \psi_s, \cos \mu_p), \quad (36c)$$

where $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ are the unit vectors pointing radially outward from the dipole center toward the observation point for the direct and reflected wave, respectively. In addition, we may regard $\hat{\mathbf{R}}_s$ as the unit vector that characterizes the direction of propagation of the ground wave as it involves the complex angle μ_p . Noting that l is the direction cosines of the dipole and using Eqs. (36a)–(36c), we can show that

$$\mathbf{l} \cdot \hat{\mathbf{R}}_1 = \cos \phi_s \cos \gamma_l + \sin \phi_s \sin \gamma_l \cos(\psi_s - \psi_l), \quad (37a)$$

$$\mathbf{l} \cdot \hat{\mathbf{R}}_2 = \cos \theta_s \cos \gamma_l + \sin \theta_s \sin \gamma_l \cos(\psi_s - \psi_l), \quad (37b)$$

and

$$\mathbf{l} \cdot \hat{\mathbf{R}}_s = \cos \mu_p \cos \gamma_l + \sin \mu_p \sin \gamma_l \cos(\psi_s - \psi_l). \quad (37c)$$

Substituting Eqs. (27), (35), and (37a)–(37c) into Eq. (13), we can recast the total sound field in a more recognizable form as

$$p = \frac{S_1}{4\pi} \left\{ (\mathbf{1} \cdot \hat{\mathbf{R}}_1) \left[\frac{1 - ikR_1}{R_1^2} \right] e^{ikR_1} + (\mathbf{1} \cdot \hat{\mathbf{R}}_2) R_p \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2} + (\mathbf{1} \cdot \hat{\mathbf{R}}_s)(1 - R_p) F(w) \times \left[\frac{1 - ikR_2}{R_2^2} \right] e^{ikR_2} \right\}. \quad (38)$$

It is clear from Eq. (38) that the total sound field, due to a randomly oriented dipole above an impedance plane, consists of a direct wave, a reflected wave, and a ground wave term as explained earlier. This asymptotic solution is expressed in a similar manner to that for a monopole source. The main difference is that each of these terms involves a factor due to the orientation of the dipole. Moreover, extra “dipole” factors, $(1 - ikR_1)/R_1^2$ and $(1 - ikR_2)/R_2^2$ are evident in Eq. (38).

Finally, it is of interest to point out that, although the sound field generated by a dipole can be obtained, in principle, by differentiating the monopole sound field with respect to the appropriate spatial coordinates, in practice, the asymptotic expression for the monopole sound field (see, for example, the asymptotic expression given in Ref. 11) is not sufficiently precise to yield satisfactory results, especially when calculating the vertical-dipole component of the sound field. It is possible, however, to carry the differentiation out in the wave-number domain and to invoke the same approximation, i.e., $\beta^2 \ll 1$, $r \gg \max(z_s, z)$ and $kR_2 \gg 1$, as above. Taherzadeh¹⁷ uses this approach and shows that it leads to the same analytic expression as given in Eq. (35). Nevertheless, we have chosen the procedure offered here since it is more straightforward and, mathematically speaking, more elegant.

II. NUMERICAL COMPARISONS

In the last section, we derived an asymptotic approximation for the total sound field due to an arbitrarily oriented dipole. As noted previously, the inverse Fourier integrals in Eqs. (17) and (18) can be evaluated by the fast-field method.¹² Predictions of the dipole form of the fast-field program (dipole-FFP) are identical to those obtained numerically for a source consisting of two adjacent out-of-phase monopoles. So the dipole-FFP may be used, in turn, to validate the asymptotic theory derived in the last section.

In the calculations presented here, the ground is characterized by a two-parameter impedance model¹⁸ given by

$$Z = 0.436(1 + i)(\sigma_e/f)^{0.5} + 19.48i\alpha_e/f, \quad (39)$$

where σ_e represents an effective flow resistivity and α_e represents an effective (exponential) rate of change of porosity with depth or an effective inverse depth. Parameter values of $\sigma_e = 38 \text{ kPa s m}^{-2}$ and $\alpha_e = 15 \text{ m}^{-1}$ are chosen since they are found to represent the apparent surface impedance of the wool felt material used in the experimental activities described later. Similar values have been found elsewhere also

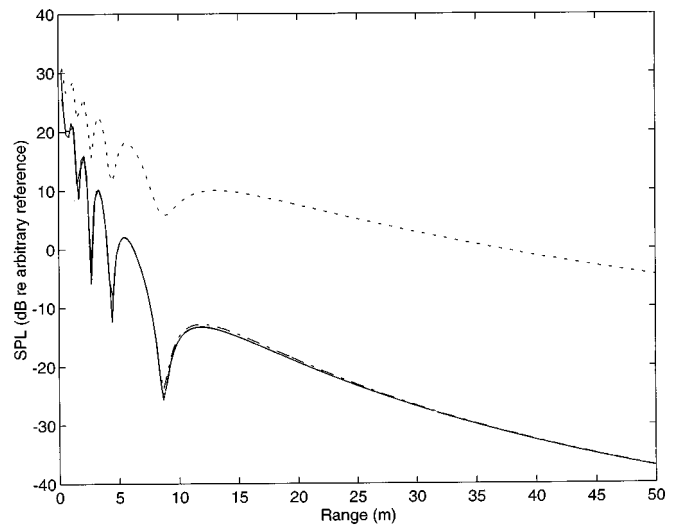


FIG. 2. Predictions of SPL from a vertical dipole at 1000 Hz above a plane characterized by a two parameter impedance model, and default parameter values 38 kPa s m^{-2} and 15 m^{-1} . Results obtained from dipole-FFP and Eq. (35) are shown by the (indistinguishable) solid lines; Generalov's solution is represented by the dash-dot line. The predicted monopole field is represented by the dotted line. The source height (z_s) is 2.0 m and the receiver height (z) is 1.0 m.

to represent the apparent the surface impedance of grassland at an airfield.¹⁹ We start by comparing the predictions of our asymptotic solution [Eq. (35)] with those obtained by Generalov⁸ and with results obtained from the dipole FFP¹² for sound level as a function of range given source and receiver at heights of 2.0 and 1.0 m, respectively, above the ground plane. Calculations are presented for two frequencies, 1 kHz and 100 Hz. Figure 2 shows predictions of the field due to a vertical dipole field as function of range at 1 kHz. The dotted line represents the prediction of the sound level due to a monopole point source at 2.0 m above the impedance plane and with the source strength normalized to give the same sound level as the dipole at 0 m range. Figure 3 is the corresponding plot at 100 Hz. Agreement between our

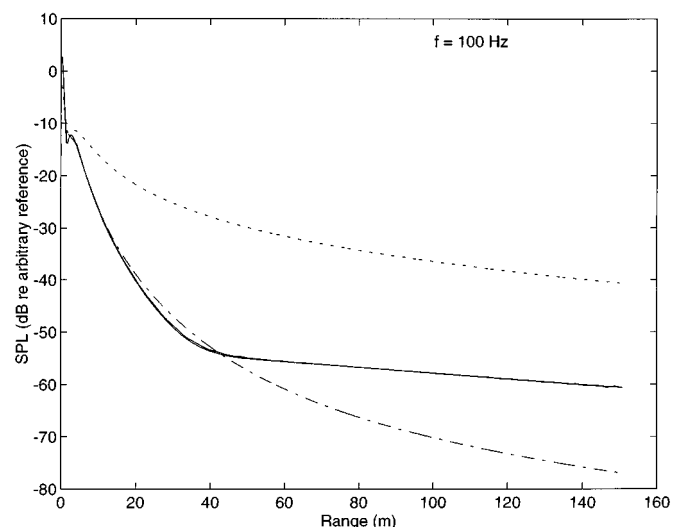


FIG. 3. Predicted SPL from a vertical dipole at 100 Hz. Geometry and key as for Fig. 2.

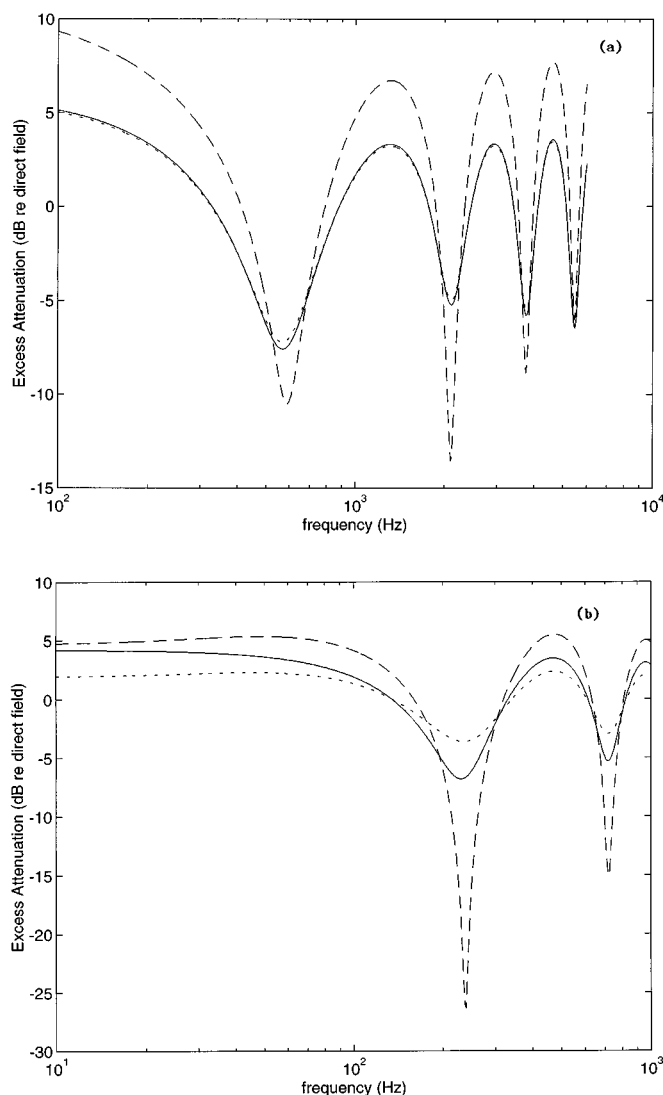


FIG. 4. Predicted excess attenuation of sound due to a monopole (solid line), horizontal dipole (dotted line), and vertical dipole (dashed line) above an impedance plane with the default parameters. The source–receiver geometry is $z_s = 1.0$ m, $z_r = 0.5$ m, and (a) range = 5.0 m, (b) range = 1.0 m.

asymptotic form [Eq. (35)] and the dipole-FFP is excellent. We remark that the calculation of the sound field based on Eq. (35) is very much faster than that due to the FFP formulation. On the other hand, at this lower frequency where the ground wave is more important, predictions obtained from Generalov's expression compare relatively poorly with those of the dipole-FFP.

Now we present comparisons between predictions for vertical and horizontal dipoles above the same impedance plane. Figure 4(a) compares predictions of excess attenuation, defined as the total sound field relative to the direct field, for monopole, vertical dipole, and horizontal dipole. The source and receiver heights are 1.0 and 0.5 m, respectively, and the separation is 5.0 m. With this geometry, there is little difference between the predicted excess attenuation spectra for monopole and horizontal dipole sources, whereas those for a vertical dipole exhibit greater interference effects. Figure 4(b) shows the same comparison but at only 1 m range. For this geometry, the angle between the horizontal

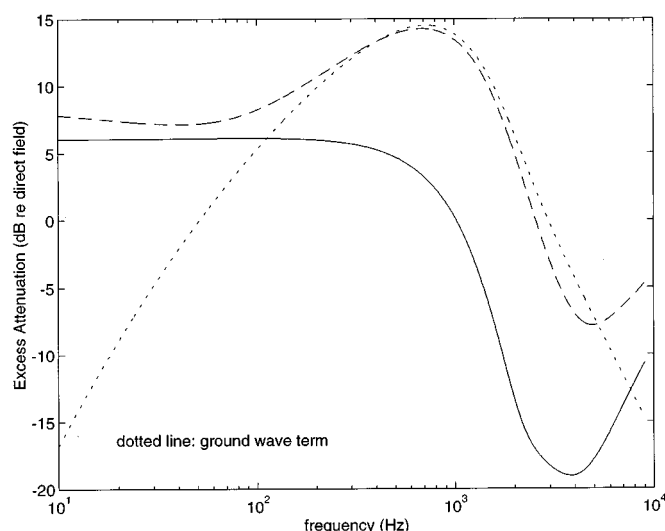


FIG. 5. Predicted excess attenuation of sound due to a monopole (solid line) and vertical dipole (dashed line) above the impedance plane. The dotted line is the ground wave component of the dipole field. The horizontal dipole spectrum for this case is indistinguishable from that for a monopole. The source–receiver geometries is $z_s = 0.1$ m, $z_r = 0.25$ m; and range = 2.0 m.

plane and the specular ray is large and so the difference between the excess attenuation spectra predicted for monopole and horizontal dipole sources is significant.

Figure 5 shows predictions of excess attenuation spectra with either monopole (continuous line) or vertical dipole (broken line) sources and receiver close to the impedance surface (0.1- and 0.25-m heights, respectively) and 2.0 m apart. Near-grazing incidence the ground wave contribution due to a vertical dipole is predicted to be particularly important. The predicted ground wave contribution to the vertical dipole field is shown by the dotted line in Fig. 5. It is worth noting that the predicted sound fields are different by interchanging the position of source and receiver. This is illustrated in Fig. 6 by excess attenuation predictions at a range of 1 m but with the source at 0.025 m and the receiver at 0.1 m over the default impedance. The difference can be explained by the fact that the polar angles joining for the direct wave for these two cases are different by π as shown in Fig. 7(a) and (b). The magnitude of the direct wave is the same for both situations but they differ by a factor of -1 . It follows that the total sound fields for both situations are not identical, see Eq. (35). Consequently, a straightforward application of the reciprocity theorem²⁰ fails for the vertical-dipole component above an impedance plane.

III. COMPARISON OF PREDICTIONS WITH DATA

Previous authors⁹ have used an un baffled loudspeaker as a dipole source. However, they have found it necessary to include substantial monopole components when comparing their numerical predictions with measurements. Anechoic chamber measurements have shown that piezoceramic disks of the type used as ultrasonic transducer elements radiate as dipoles near their resonant frequencies with the main lobes normal to the disk faces. An example radiation pattern measured at a radius of 0.5 m is shown in Fig. 8. Although the use of these compact essentially monofrequency sources re-

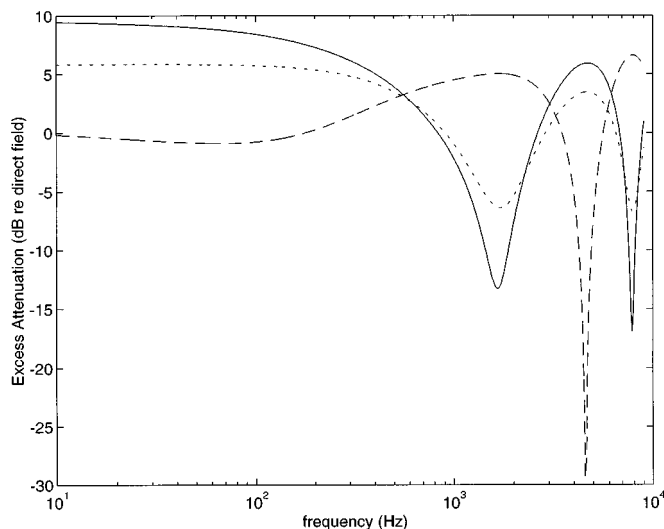


FIG. 6. Predicted excess attenuation of sound due to a monopole (solid line) and vertical dipole (dotted line) with $z_s=0.25$ m and $z=0.1$ m and range=1.0 m. The dashed line represents the prediction for a vertical dipole with $z_s=0.1$ m and $z=0.25$ m (dashed line). In the latter case, the direct and the reflected "rays" are out of phase and cancel at lower frequencies.

stricts validation of predictions to variation with range, the dipole behavior is clear and assured at relatively small ranges thus making it straightforward to make measurements in a confined space. A typical disk diameter is 0.02 m. Consequently dipole behavior was assured at distances greater than 0.1 m. Each of two piezoceramic transducer disks, with measured resonance frequencies of 2915 and 4069 Hz, respectively, were suspended 0.03 m above an impedance floor in an anechoic chamber of wedge-tip to wedge-tip internal dimensions $3 \times 3 \times 3$ m. Each disk, with a diameter of 2 cm, was supported on a steel rod clamped to a tripod stand 1.0 m

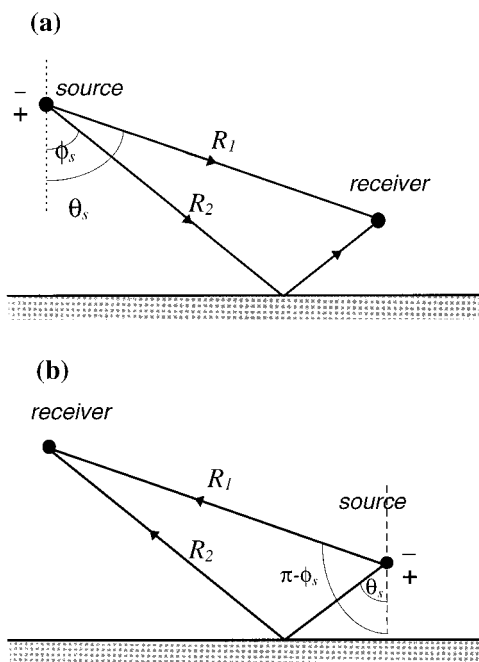


FIG. 7. Schematic diagram to demonstrate the difference in the polar angles when the positions of source and receiver are interchanged.

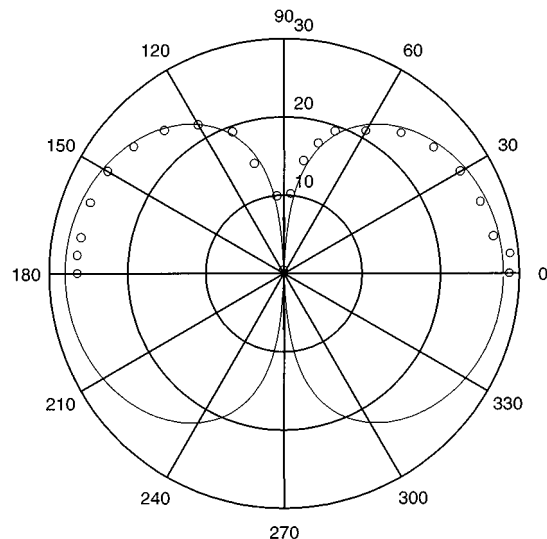


FIG. 8. The directivity pattern of a transducer disk (circles) compared with the directivity of a dipole (solid line). The disk was positioned in a plane perpendicular to the plane of the paper and parallel to the 90° axis.

away. A Bruel & Kjaer type 4311 1.3-cm-diam condenser microphone, fitted with a preamplifier and suspended with light flexible string at 0.02-m height and at several distances between 0.02 and 1.5 m from the source, was used as the receiver for all of the experiments. The height of the dipole source is measured from the center of the piezoceramic disk. The choice of the maximum distance was determined by the size of the hardwood board and the space available in the anechoic chamber. A 0.016-m-thick layer of wool felt placed on a varnished hardwood board measuring $1.2 \times 1.2 \times 0.02$ m (thick) was used as the impedance floor. A reference measurement, without the impedance floor, was used as the direct field in subsequent processing. A PC-based Maximum Length Sequence System Analyzer (MLSSA)²¹ was used both as the signal generator for the source and as the analyzer for subsequent signal processing. The MLSSA system controls the source output to be in the form of a pseudorandom sequence of pulses. The microphone signal is analyzed with respect to the known output sequence thus eliminating background noise effects. After processing and FFT, each spectrum level was normalized by the direct field measured earlier. All the steps were carried out with MLSSA software, the final output being the desired excess attenuation spectrum.

To characterize the impedance of the floor, separate measurements of the excess attenuation were made using a point source consisting of a Tannoy driver fitted with a tube 15-cm-long and 3-cm internal diameter. The source and receiver were suspended at 0.05- and 0.1-m heights and 2-m separation above the felt. These measurements were subsequently fitted by means of a two-parameter impedance model¹⁸ described previously. Best fit parameters were found to be 38 kPa s m^{-2} and 15 m^{-1} . Examples of the resulting measured and fitted excess attenuation measurements are shown in Fig. 9. Substituting these values into Eq. (39), the real and imaginary parts of the impedance can be determined as (1.57, 1.67) and (1.33, 1.40) at 2915 and 4096 Hz, respec-

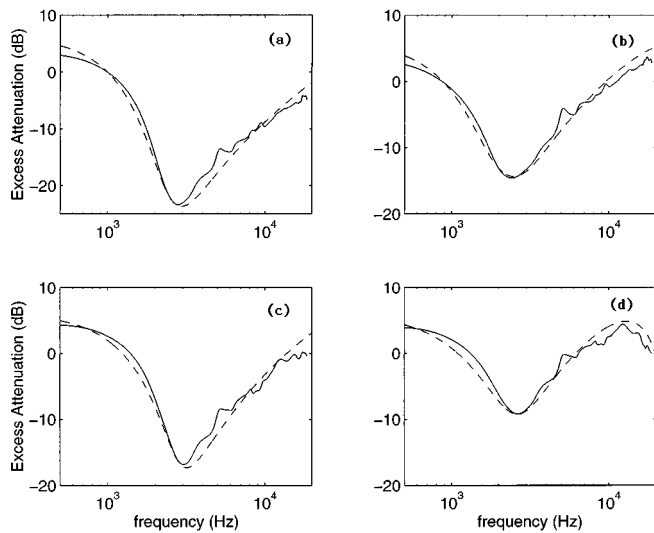


FIG. 9. Measured (solid line) and fitted (broken line) excess attenuation spectra due to a monopole source above a rigid-backed 1.6-cm-thick felt: (a) $z_s = z = 0.05$ m, range = 2.0 m. (b) $z_s = 0.05$ m, $z = 0.15$ m, range = 2.0 m. (c) $z_s = z = 0.05$ m, range = 1.0 m. (d) $z_s = 0.05$ m, $z = 0.15$ m, range = 1.0 m. In all four cases the ground parameters are 38 kPa s m^{-2} and 15 m^{-1} .

tively. These parameters have been used together with the measurement geometries to obtain the predictions of dipole propagation above an impedance plane for comparison with measured data.

Figure 10 shows measured and predicted total sound fields at 4096 Hz as a function of range with the source disk axis horizontal and at a height of 0.03 m above the felt surface. The agreement between data and predictions according to Eq. (29) is good. Also shown is the prediction for the sound field due to a monopole above the felt surface normalized to the same level at the first measurement point (0.1 m). Corresponding results for the horizontal dipole at 2915 Hz are not included in the present paper for brevity. Neverthe-

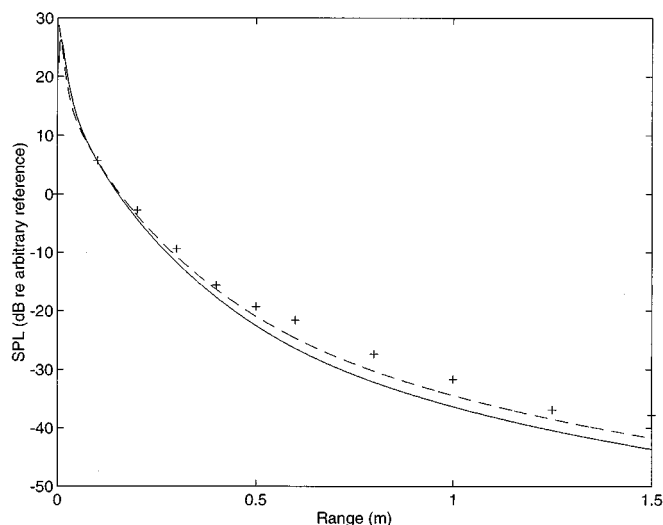


FIG. 10. Measured (crosses) and predicted sound fields from a horizontal dipole at 4096 Hz with $z_s = 0.03$ m and $z = 0.02$ m. Asymptotic form (solid line) and monopole field (dashed line) normalized to coincide at the first measured point (0.1 m in this case).

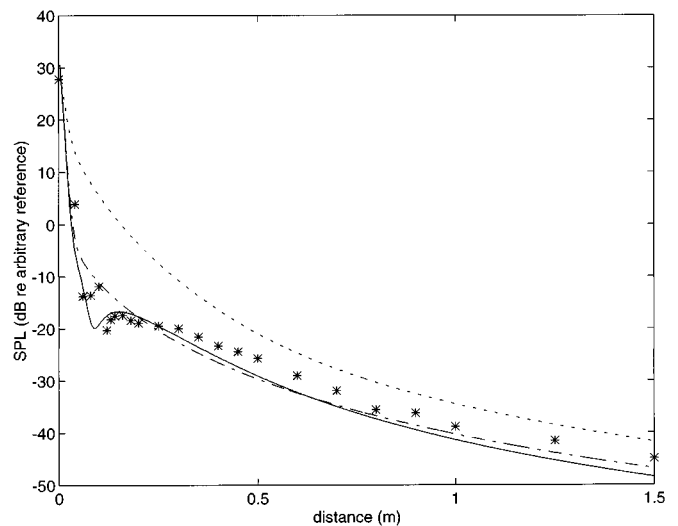


FIG. 11. Measured (asterisks) and predicted sound field from a vertical dipole at 4096 Hz with $z_s = 0.03$ m and $z = 0.02$ m. Asymptotic form (solid line); Generalov's solution (dash-dot line), and monopole field (dotted line) normalized to coincide at the first measured point (0.0 m in this case).

less the results are, as expected, rather similar to that of Fig. 10. Figure 11 shows measured data with the loudspeaker disk axis vertical and predictions according to Eq. (35). Again the agreement between data and predictions is good. Also shown are the predictions of the theory due to Generalov⁸ and for the sound level due to a monopole normalized as before. Corresponding results for the vertical dipole using the 2915 Hz source at a height of 0.035 m are shown in Fig. 12. There are significant differences between monopole and dipole fields above an impedance plane at short range. Moreover the Generalov theory fails to predict the measured interference at short range for either frequency.

IV. CONCLUSIONS

A closed-form solution for an arbitrarily orientated dipole above an impedance plane has been derived.

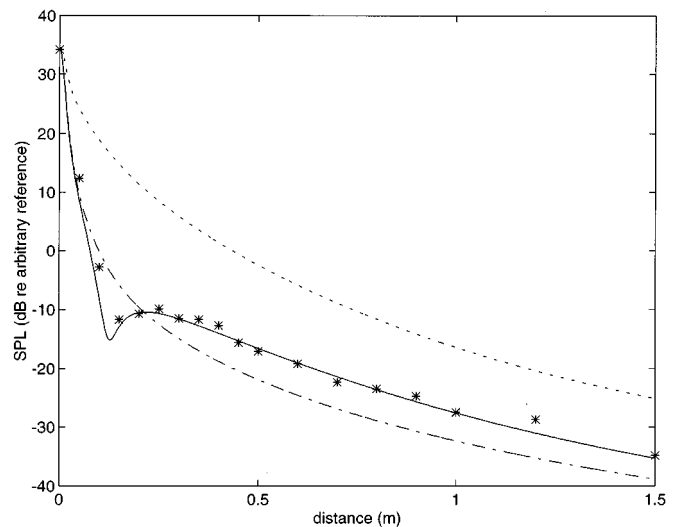


FIG. 12. Measured (asterisks) and predicted sound field from a vertical dipole at 2915 Hz with $z_s = 0.03$ m and $z = 0.02$ m. Asymptotic form (solid line); Generalov's solution (dash-dot line), and monopole field (dotted line) normalized to coincide at the first measured point (0.0 m in this case).

Asymptotic approximations have been obtained for horizontal and vertical dipoles which have the same form as the classical result for a monopole thus permitting a similar physical interpretation. Comparisons with numerical integration results and with experimental data have validated the asymptotic formulas and have shown their superiority over those obtained by Generalov.⁸ Specifically, it has been shown that Generalov's approximation does not predict the ground wave component accurately.

The ground wave component is predicted to be particularly significant for a vertical dipole source above an impedance plane. This may be understood as a consequence of the preferred downward radiation from the source. Moreover the sound field above an impedance plane due to a vertical dipole is not reciprocal. The difference between the field due to a horizontal dipole and a monopole has been found to be slight near-grazing incidence above an impedance plane. However, when the angle between the horizontal and the specularly reflected ray is large, significant differences between sound fields due to horizontal dipole and monopole are predicted.

Good agreement with experimental data for propagation from dipole sources comprised of piezoceramic disks has been obtained without including arbitrary monopole corrections such as those found necessary for agreement with previously published data.⁹

ACKNOWLEDGMENTS

This work was supported by the EPSRC (UK) through Grant Ref. No. GR/J42052. The authors are grateful to Mrs. Qiang Wang for assistance with the impedance characterization and to Dr. David Hothersall for suggesting the use of piezoceramic transducer disks as dipole sources.

¹K. Attenborough, S. I. Hayek, and J. M. Lawther, "Propagation of sound over a porous half-space," *J. Acoust. Soc. Am.* **65**, 1493–1501 (1980).

²P. M. Morse, *Vibration and Sound* (Acoustical Society of America, New York, 1983), pp. 318–319.

³S. Peters, "The prediction of railway noise profiles," *J. Sound Vib.*, 32–87 (1974).

⁴M. E. Goldstein, *Aeroacoustics* (McGraw-Hill, New York, 1976).

⁵B. Brouard, D. Lafarge, and J. F. Allard, "Measurement and prediction of the reflection coefficient of porous layers at oblique incidence and for inhomogeneous waves," *J. Acoust. Soc. Am.* **99**, 100–107 (1996).

⁶A. Banos, *Dipole Radiation in the Presence of a Conducting Half-space* (Pergamon, Oxford, 1966).

⁷D. S. Jones, *Acoustics and Electromagnetic Waves* (Clarendon, Oxford, 1986).

⁸A. V. Generalov, "Sound field of a multipole source of order N near a locally reacting surface," *Sov. Phys. Acoust.* **33**, 492–496 (1987).

⁹Z. Hu and J. S. Bolton, "Sound propagation from an arbitrarily oriented multi-pole placed near a plane, finite impedance surface," *J. Sound Vib.* **170**, 637–66 (1994).

¹⁰L. M. Brekhovskikh, *Waves in Layered Media* (Academic, London, 1980).

¹¹C. F. Chien and W. W. Soroka, "Sound Propagation along an impedance plane," *J. Sound Vib.* **81**, 413–424 (1975).

¹²S. Taherzadeh and K. Attenborough, "Sound propagation from dipole and monopole sources in a stratified fluid above a layered poro-elastic solid," *Computational Acoustics and its Environmental Applications*, edited by C. A. Brebbia (Computational Mechanics, Southampton, 1995), pp. 45–54.

¹³The similar form can be found in, for example, Ref. 7, pp. 348–351. Note also that the time dependent factor, $e^{i\omega t}$, is used in Ref. 7 whereas $e^{-i\omega t}$ is understood in the present paper.

¹⁴A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1989), pp. 165–171.

¹⁵M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. (Dover, New York, 1972).

¹⁶G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1944), pp. 196–198. [The given formula is obtained by substituting $u = y^2/2$ and $v = n$ where $n \geq 1$.]

¹⁷S. Taherzadeh, "Sound propagation in inhomogeneous media," Ph.D. thesis, The Open University, 1996.

¹⁸K. Attenborough, "Ground parameter information for propagation modeling," *J. Acoust. Soc. Am.* **92**, 418–427 (1992).

¹⁹K. Attenborough, K. M. Li, and S. Taherzadeh, "Propagation from a broad-band source over grassland: Comparison of data and models," *Proc. Inter-Noise 95* **1**, 319–322 (1995).

²⁰J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 2, sec. 294.

²¹D. D. Rife and J. Van der Kooy, "Transfer-function measurement with Maximum-Length Sequences," *J. Audio Eng. Soc.* **37**(6), 419–443 (1989).

Oscillation modes of supersonic multijets

Y. Umeda and R. Ishii

Division of Aeronautics and Astronautics, Department of Engineering Science, Kyoto University, Kyoto, 606 Japan

(Received 2 October 1995; revised 20 September 1996; accepted 20 December 1996)

In this paper, oscillation modes of single jet, twin jet, and multijet with square configuration were investigated by using optical and acoustical observations. In the case of a single jet, axisymmetrical, helical, and lateral oscillation modes appeared with increasing pressure ratio of the jet. In the case of twin jets, two counter rotating helical oscillation modes and reversely lateral oscillation modes along a line through the center line of the two jets were observed for two different pressure ratios. In the case of multijets with square configuration, the quadrupolelike oscillation mode, four counter rotating helical oscillation modes, and a symmetrical oscillation mode were observed with increasing pressure ratio. The first and the last oscillation modes of the multijet with square configuration are unexpected based on observations for a single jet. © 1997 Acoustical Society of America. [S0001-4966(97)00505-5]

PACS numbers: 43.28.Ra, 43.50.Nm, 43.28.Py [LCS]

INTRODUCTION

In a supersonic free jet issuing from a circular convergent nozzle, there are several shock cell structures and the large coherent vortical structures are convected downstream and interact with the shock waves at the rear edge of the shock cells. The interaction between the vortical structure and the shock wave is very effective to produce strong acoustic pulses. The resulting sound wave propagates upstream and stimulates the thin shear layer near the nozzle lip. In such a manner, we have a well-known feedback loop proposed by Powell (1953a).

In measured spectra of sound emitted from a supersonic jet, we can see a broadband frequency noise component as well as very strong narrow-band noise, called a screech tone, radiated by the feedback loop instability. This strong feedback loop instability is responsible for the self-sustained oscillation of the jet. At first glance, such a feedback instability might seem relatively simple, but actually is quite complex. In the frequency characteristics of screech tones radiated from a circular underexpanded jet for a thick-lip nozzle, we can see several discontinuities. These discontinuities suggest that the circular jet oscillates in different oscillation modes in the different pressure ratio ranges.

Some investigations about the instability of supersonic jets have been performed by several researchers. Westley and Wooley (1975) found that a circular jet oscillates in the helical mode at the pressure ratio $R=4.7$. Davies and Oldfield (1962) found that five stages exist in the pressure ratio range less than about 6.0 and the first four oscillation modes are axisymmetric, axisymmetric, lateral, and helical, respectively, with increasing pressure ratio. Recently, Powell *et al.* (1990, 1992) showed that the last oscillation mode at highest pressure ratio R (mode D) is lateral. They also showed that the plane of oscillation of the last lateral mode (mode D) rotates, vibrates, and dwells, while the first lateral mode (mode B) occasionally changes to a helical mode at the same frequency. Yu and Seiner (1983) observed that when the convergent-divergent (CD) nozzle of Mach 2 design was

used in an overexpanded mode, a helical mode was obtained.

For a rectangular jet of aspect ratio $AR=16.9$, Powell (1953b) found that the frequencies of the screech tones decrease regularly with increasing pressure ratio R . Hammitt (1961) showed that the oscillation mode of the rectangular jet is lateral. Powell and Kozu (1991) reported that for jets with $AR=4.8$, two modes exist, the first of which is symmetric ($R<2.16$). Suda *et al.* (1993) showed that symmetric and lateral modes exist for the jet issuing from rectangular CD nozzles of $AR=8$ and 10 with side walls. Merle (1956) reported that two modes exist in the case of a square jet. Umeda and Ishii (1994) showed that there are only two oscillation modes for underexpanded jets exhausting from square and equilateral triangular nozzles, and these are axisymmetric and lateral modes with increasing jet pressure ratio.

The propulsion systems of many aircraft and rockets use two or more circular jets. When multiple supersonic jets are closely spaced, the sound level of the screech can be greatly amplified due to a coupling process between the jets. Thus, the instability of multiple interacting supersonic jets has generated much interest in recent years. So far, the noise and coupling process between two closely spaced jets has been studied by a few researchers. Seiner *et al.* (1988) showed that interacting jets can generate dynamic pressures exceeding the fatigue failure limit for aircraft structures. Wlezien (1989) found that for closely spaced nozzles, coupling occurs at low jet Mach numbers and is suppressed at high Mach numbers. Shaw (1990) found that the use of tabs and secondary jets was effective to reduce the screech tone.

The objective of this experimental study is to investigate the oscillation modes of round jets for a pair and a group of four with a square configuration of their center lines.

I. EXPERIMENTAL APPARATUS

The schematic views of the nozzle configurations used in this experiment are shown in Fig. 1. The cylindrical nozzles were attached to a 100-mm-diam flange to the set-

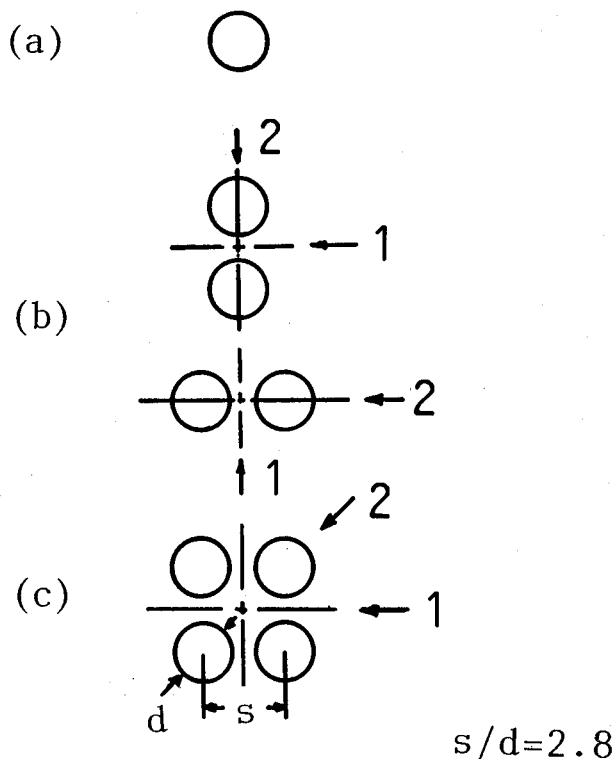


FIG. 1. Nozzle configuration.

ting chamber. The nozzles had an exit internal diameter of $d = 5$ mm. The length of the nozzles was 50 mm. The nozzle lip thickness was 0.5 mm and the center-to-center spacing of the nozzles was fixed at $S = 14$ mm ($S/d = 2.8$). In these figures, arrows 1 and 2 show the directions of optical observation. The flange was covered by a dome-shaped urethane foam in order to reduce the reflection of sound waves.

All the experiments were carried out within a simplified anechoic chamber. First, the frequency characteristics of the screech was measured in order to find the pressure ratios where the discontinuities occur. Next, optical and acoustical observations were performed at various pressure ratios associated with individual oscillation modes in order to discriminate the jet oscillation mode.

All of the signal outputs from the two microphones were analyzed on a Yokogawa AR-1100 analyzing recorder. In the optical observation, a conventional single pass schlieren system was used with a Sugawara type MS-230 spark light source of duration about $1 \mu\text{s}$.

II. EXPERIMENTAL RESULTS

A. Frequency characteristics of screech tone

The frequency characteristics of the screech tones radiated by three jet configurations as shown in Fig. 1 at various pressure ratios R from 2.00 to 6.33 were measured by a microphone placed in the near-sound field. These pressure ratios correspond to ideally expanded jet Mach numbers of 1.05 and 1.86, respectively.

Figure 2 shows the experimental results for a thin-lip nozzle. In these figures only the fundamental frequencies were plotted. The solid circles indicate dominant tones, de-

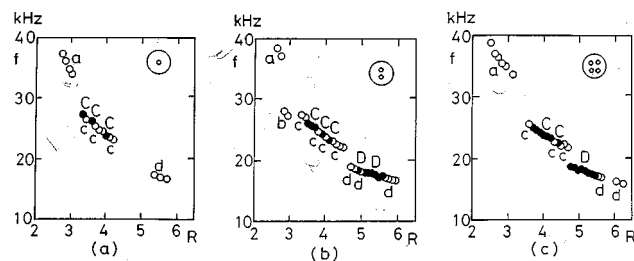


FIG. 2. Frequency characteristics of screech tone; (a) single jet, (b) twin jet, and (c) multijet with square configuration.

finied here as those that exceed the local broadband noise by at least 10 dB, usually by more than 20 dB and are labeled with capital letters C and D. The open circles mark secondary tones that exceed the broadband noise by 5–10 dB and are labeled with lowercase letters a, b, c, and d.

Obviously, the steady decrease of the screech frequency with increasing pressure ratio is interrupted by two or three frequency jumps. As shown in Fig. 2, the frequency characteristics of screech tones radiated from single jets, twin jets, and multijets with square configuration are only slightly different from each other.

B. Optical observation

In order to investigate the oscillation modes of jets, first the schlieren photographs with an exposure of $1 \mu\text{s}$ were taken to display the jet configuration and the associated near-sound field for each of the dominant modes. In this visualization, all the jet configurations except for single jets were observed from two different directions.

1. Single jet

Figure 3 shows the flow field and near-sound field of the single jets. These jets belong to three different stages a, C, and d, respectively. Sound waves for stage a are symmetrical. But in the case of stages C and d, they are out of phase across the jet.

2. Twin jet

Figure 4(a) and (c) shows the photographs of the twin jet which were observed from direction 1 shown by the arrow in Fig. 1 (a line of symmetry between jets). Figure 4(b) and (d) shows the corresponding photographs which were observed from the direction of arrow 2 (a line through the center line of a pair of jets). These jets belong to stages C and D, respectively.

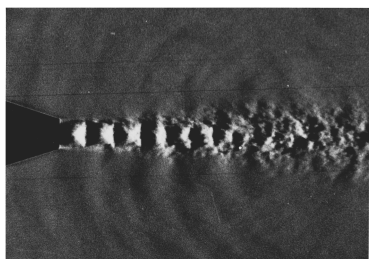
3. Multijet with square configuration

Figure 5(a), (c), and (e) shows the photographs of flow field and near-sound field of the multijet with square configuration observed from direction 1. Figure 5(b), (d), and (f) shows the corresponding photographs which were seen from direction 2. These jets belong to stages a, C, and D, respectively.



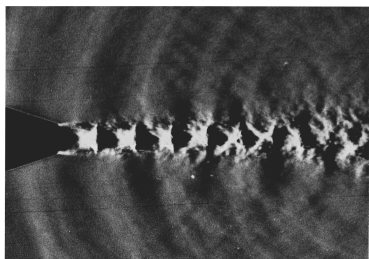
R=3.03 stage a

(a)



R=3.90 stage C

(b)



R=5.35 stage d

(c)

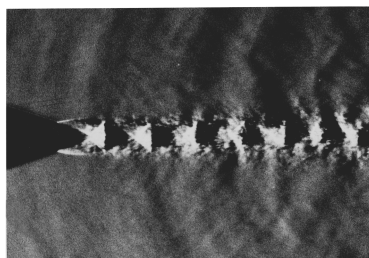


FIG. 3. Schlieren photographs of single jet.

C. Acoustical observation

From only the above-mentioned optical observations, the oscillation modes of jets cannot be discriminated. In order to discriminate the oscillation modes of these jets, it is necessary to get information on the three-dimensional structure of wavefronts radiated from the jets. The information can be obtained by measuring the phase differences of the screech tone radiated by the jets as shown in the previous investigations by Powell *et al.* (1990, 1992) and Umeda and Ishii (1994). The phase differences between the sound signals received by the two microphones were measured by a correlation method on a digital analyzer. This phase difference between the signals will correspond to a particular three-dimensional structure of wavefronts as shown by Powell *et al.* (1990, 1992) and Umeda and Ishii (1994).

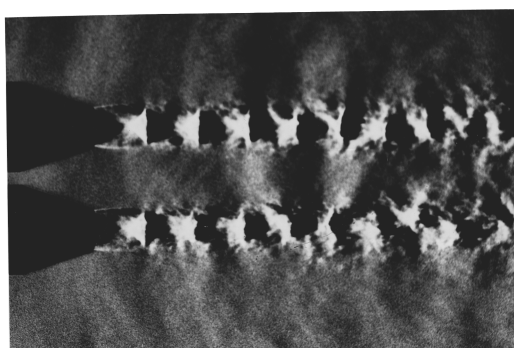
1. Single jet

Applying this technique to the single jet issuing from a thin lip nozzle, the relationship between phase difference ψ and separation angle θ of the microphones were obtained and are shown in Fig. 6(a)–(c). These correspond to the modes a, C, and d, respectively. The open circles show the average value of the eight data sets and the bar represents the scatter.

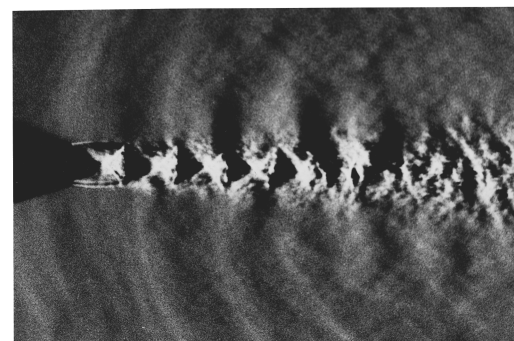
For the oscillation mode of stage a (mode a) ($R=3.03$), the data are close to zero [Fig. 6(a)]. This means that the mode a is axisymmetric or toroidal. For the mode of stage C (mode C) ($R=3.90$), the linear relationship between

R=3.71 stage C

(a)

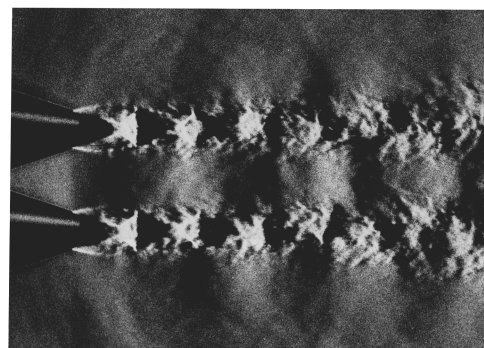


(b)



R=5.16 stage D

(c)



(d)

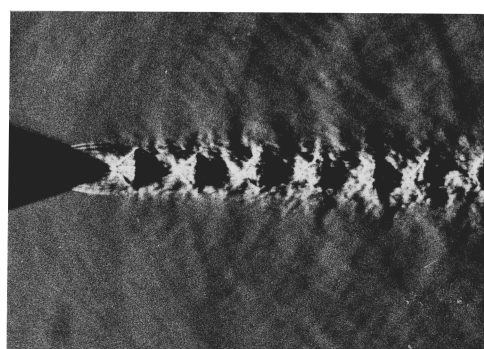


FIG. 4. Schlieren photographs of twin jet.

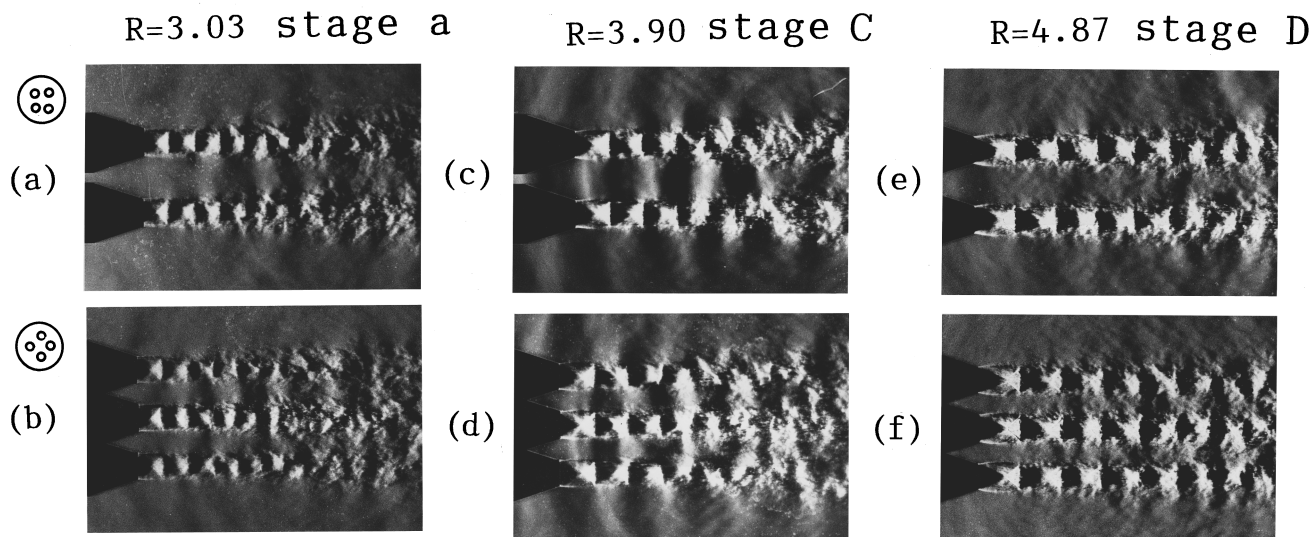


FIG. 5. Schlieren photographs of multijet with square configuration.

the phase difference ψ and the microphone separation θ indicates a helical mode [Fig. 6(b)]. For the mode d ($R = 5.35$), there is the clustering about the values $\psi=0$ and π , showing the lateral mode [Fig. 6(c)]. These results for

single jets completely agree with the results shown by Powell *et al.* (1990, 1992).

In the figures on the right side, the axisymmetrical, helical, and lateral oscillations are shown by a circle, a circle with arrows, and a linear arrow, respectively. The positions of the fixed microphone M_f and the microphone separation angle θ are shown in Fig. 6(c).

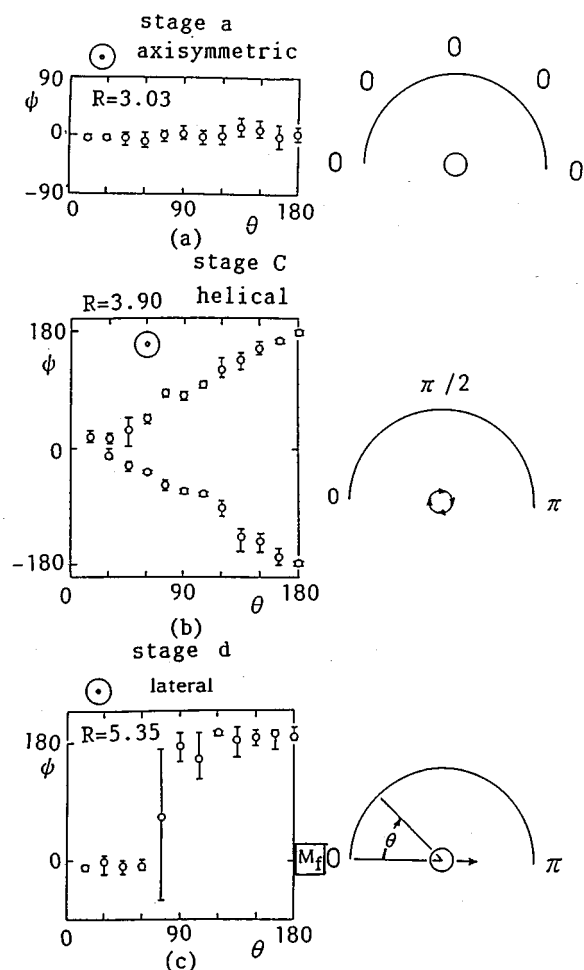


FIG. 6. Relationship between phase difference ϕ and separation angle θ of microphones for single jets.

2. Twin jet

In the case of a twin jet, the relationship between the phase difference ψ and the microphone separation angle θ were measured at two different positions of fixed microphone M_f for the nozzle configurations shown in Fig. 7. The experimental results obtained for the stage C (mode C) ($R = 3.71$) are shown in Fig. 7(a) and (b). Since the single jet oscillates in the helical mode at this pressure ratio, these experimental data will be reasonably understood by assuming the two counter rotating helical oscillations as shown in Fig. 7(a) and (b).

For the mode D ($R = 5.16$), the relationship between the phase difference ψ and separation angle θ of the microphones are shown in Fig. 7(c) and (d). Since the single jet oscillates in the lateral mode at this pressure ratio, these experimental data will also be reasonably understood by assuming the presence of the two lateral oscillations along a line through the center line of the two nozzles as shown in Fig. 7(c) and (d). So, in this case, the twin jet oscillates in a reversely lateral oscillation mode along a line through the center line of two nozzles.

3. Multijet with square configuration

For the multijet with square configuration of their center line, the relationship between the phase difference ψ and separation angle θ of the microphones are shown in Fig. 8(a)–(c). In this case, the data were also obtained when the fixed microphone M_f was placed at the angular position $\theta=0^\circ$ as shown in Fig. 6(c).

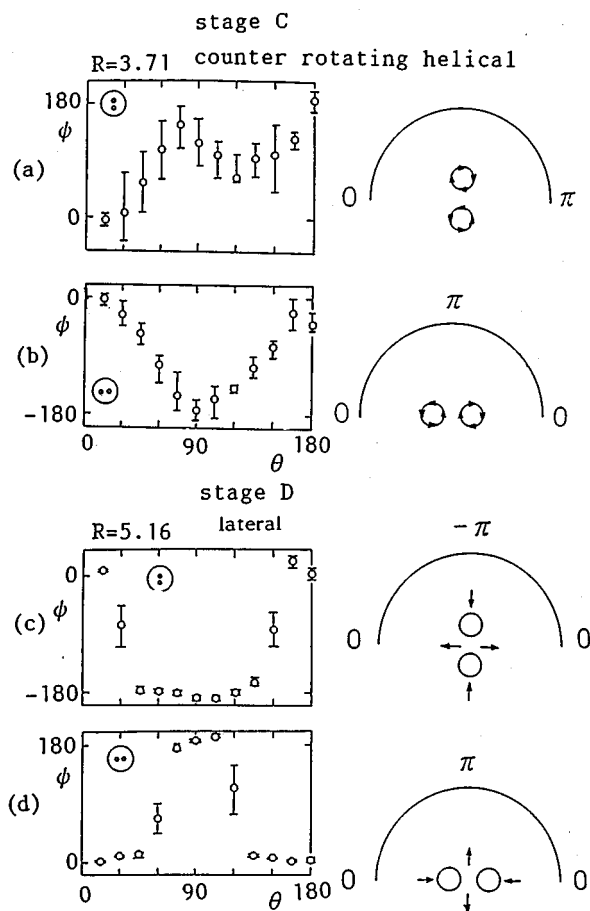


FIG. 7. Relationship between phase difference ψ and separation angle θ of microphones for twin jets.

Figure 8(a) shows the result for the mode of stage a (mode a) ($R = 3.03$). These data show that $\psi = 0$ at separation angles about $\theta = 0^\circ$, 90° , and 180° , and $\psi = \pm\pi$ at each intermediate separation angle. Although the single jet oscillates in the axisymmetric mode at this pressure ratio, these data show that four jets may oscillate in lateral mode along the diagonal lines of square nozzle configuration, namely, in a quadrupolelike mode.

Figure 8(b) shows the result for the mode C ($R = 3.90$). Since the single circular jet oscillates in a helical mode at this pressure ratio, these experimental data will reasonably suggest that four counter rotating helical oscillation modes occur.

Figure 8(c) shows the result for the mode D ($R = 4.87$). Although the single circular jet oscillates laterally at this pressure ratio, these experimental data show that four jets oscillate in a symmetrical mode.

III. SOUND-PRESSURE LEVELS

For three different nozzle configurations, the variation of the sound pressure levels (SPL) of the screech tones with pressure ratio R of the jets were measured by the fixed microphone M_f located in the near-sound field, and are shown in Fig. 9(a)–(c). For the twin jet and the multijet with square configuration, the sound-pressure levels were measured at

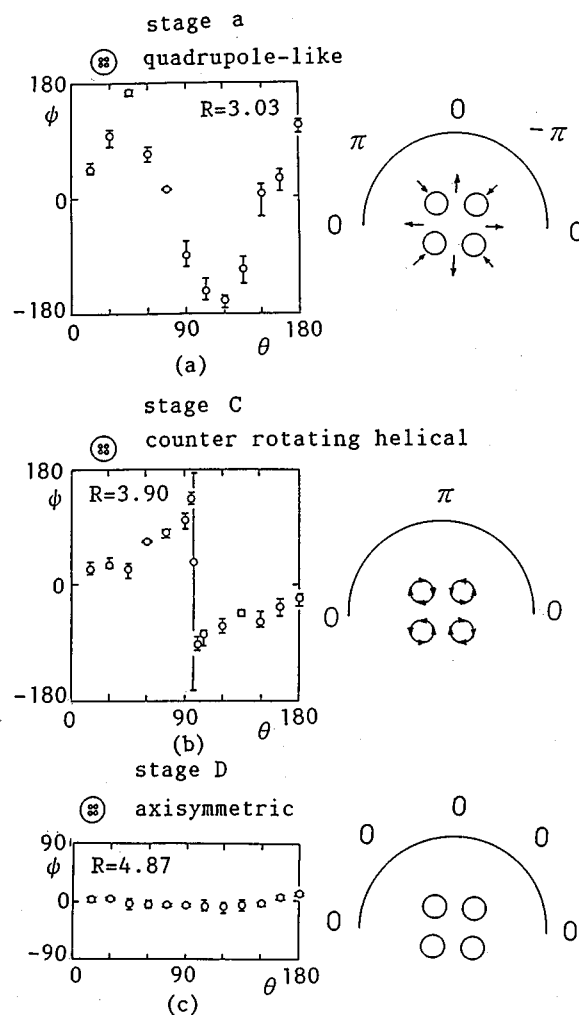


FIG. 8. Relationship between phase difference ψ and separation angle θ of microphones for multijets with square configuration.

two different microphone positions relative to the nozzle system. The different symbols in these figures correspond to the different oscillation modes.

These experimental results show an interesting fact that although the SPL for the single jet increases slightly with increasing pressure ratio R , the maximum sound-pressure level is reached at the pressure ratio R of about 5 (mode D) for a twin jet and four jets in a square configuration. The differences between the maximum sound-pressure levels of these multiple jets and that of a single jet are about 20 and 10 dB, respectively.

IV. OVERALL SOUND-PRESSURE LEVEL

Figure 10 shows the overall sound-pressure levels (OASPL) for these nozzle configurations measured in the far-sound field. From this figure, it can be seen that the OASPL for a twin-jet and four-jet configuration differs from that for a single jet by about 3 and 6 dB, respectively, at the same pressure ratios. This result is reasonable, because Fig. 10 shows that these OASPL are proportional to the cross sectional area of the nozzles.

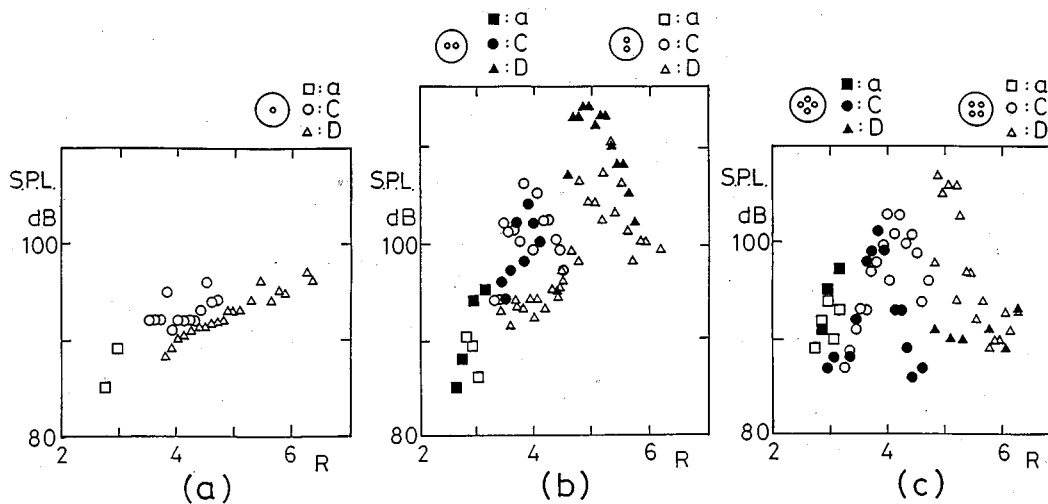


FIG. 9. Variation of sound-pressure level of screech tone with pressure ratio (near-sound field); (a) single jets, (b) twin jets, and (c) multijets with square configuration.

V. DIRECTIONALITY OF SOUND WAVES

The directionalities of the sound-pressure levels were measured by traversing a microphone in the horizontal plane. Figure 11 shows the directionalities of sound waves radiated from the single jet at three pressure ratios. The directionalities of the OASPL are shown in Fig. 11(a) and those of the SPL for discrete frequencies are shown in Fig. 11(b), respectively. In these figures, the letter α represents the angle measured from the jet axis. Obviously only one dip is observed in each directionality curve for the angle from about $\alpha = 65^\circ$ to 100° corresponding to the pressure ratios. The angle of dip in the directionality curve of the OASPL decreases with increasing pressure ratio.

In the case of a twin jet and multijet with square configuration, the directionalities of sound waves were measured

at two different inclination angles of a line through the center line of the twin jet and a diagonal line of the multijet with square configuration to a traversing plane of microphone (horizontal plane) as shown in Figs. 12 and 13. The directionality of the sound waves radiated from the twin jet at two pressure ratios are shown in Fig. 12. The OASPL at two inclination angles of the twin jet are very similar to each other at each pressure ratio and have only one dip. But, the directionality of the SPL at the pressure ratio $R=3.71$ is complicated when the two nozzles were placed in a horizontal plane as shown by the solid circles in Fig. 12(b). The directionalities of sound waves radiated from the multijet with square configuration at three pressure ratios are shown in Fig. 13. The OASPL at two inclination angles of the diagonal lines are also very similar to each other. The SPL curve for the pressure ratio $R=3.90$ (mode C) has one deep dip and that for the pressure ratio $R=4.87$ (mode D) has one shallow dip. Whereas, the directionality curve of the SPL for the pressure ratio $R=3.03$ [mode a, Fig. 13(b)] has two dips. This somewhat complicated directionality for the pressure ratio $R=3.03$ may be attributed to the quadrupolelike oscillation mode of the jet.

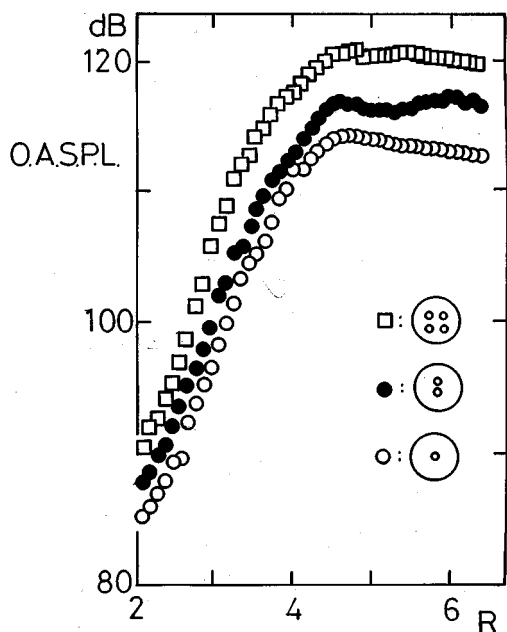


FIG. 10. Comparison of overall sound-pressure levels (OASPL) for three nozzle configurations (far-sound field).

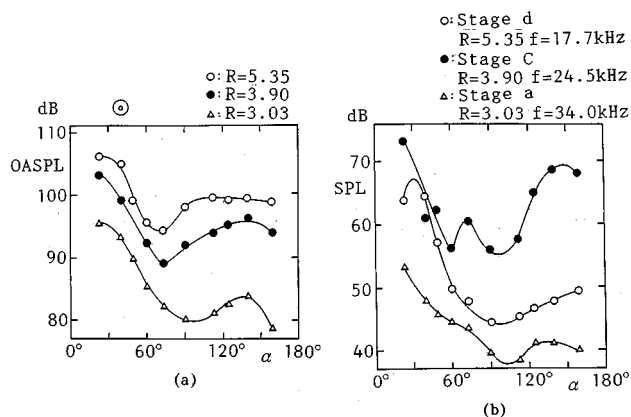


FIG. 11. Directionality of sound radiated from single jets; (a) OASPL at pressure ratios $R=3.03$, 3.90, and 5.35, (b) SPL for discrete frequencies at pressure ratios $R=3.03$, 3.90, and 5.35.

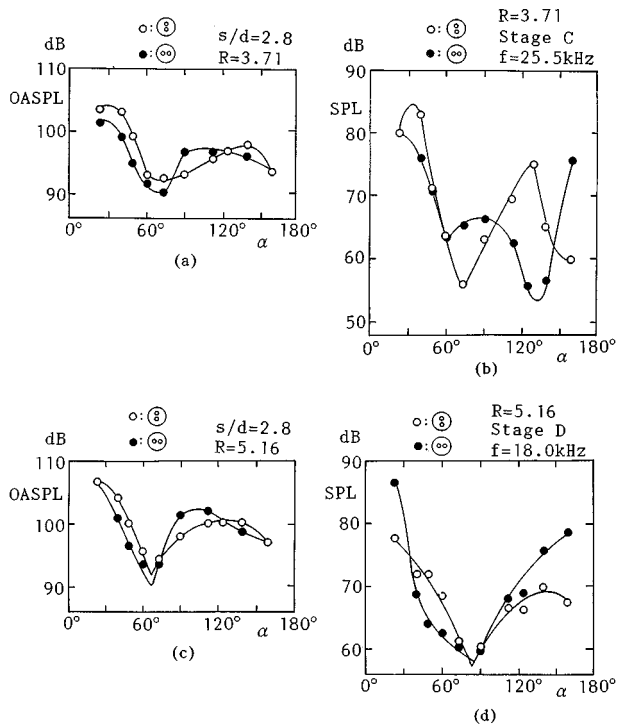


FIG. 12. Directionality of sound radiated from twin jets; (a) OASPL at pressure ratio $R=3.71$, (b) SPL for discrete frequency at pressure ratio $R=3.71$, (c) OASPL at pressure ratio $R=5.16$, and (d) SPL for discrete frequency at pressure ratio $R=5.16$.

VI. DISCUSSION AND CONCLUSIONS

In this experimental investigation, the oscillation mode of a supersonic circular single jet, twin jet, and multijet with square configuration were examined using optical and acoustical observations.

In the case of a single jet, axisymmetrical, helical, and lateral oscillation modes appear with increasing pressure ratio of the jet. In the case of a twin jet, two counter rotating helical oscillation modes and two reversely lateral oscillation modes along a line through the center line of two jets are observed for the two different pressure ratios investigated. In the case of a multijet with square configuration, the quadrupolelike oscillation mode where four jets oscillate laterally along the diagonal lines of square, four counter rotating helical oscillation modes and symmetrical oscillation mode appear with increasing pressure ratio of the jet.

The experimental results for the single jet and twin jet are quite reasonable. Because, the results for a single jet completely agree with the results shown by Powell *et al.* (1990, 1992) and the results for the twin jet correspond to those obtained for the single jet.

But, two of the three results for the multijet with square configuration are unexpected and interesting. Namely, for the mode a, although the single jet oscillates in an axisymmetric mode, the four jets in a square configuration oscillate in a quadrupolelike oscillation mode, and for the mode D, although the single jet oscillates in a lateral mode, the four-jet configuration oscillates in a symmetric mode. These results

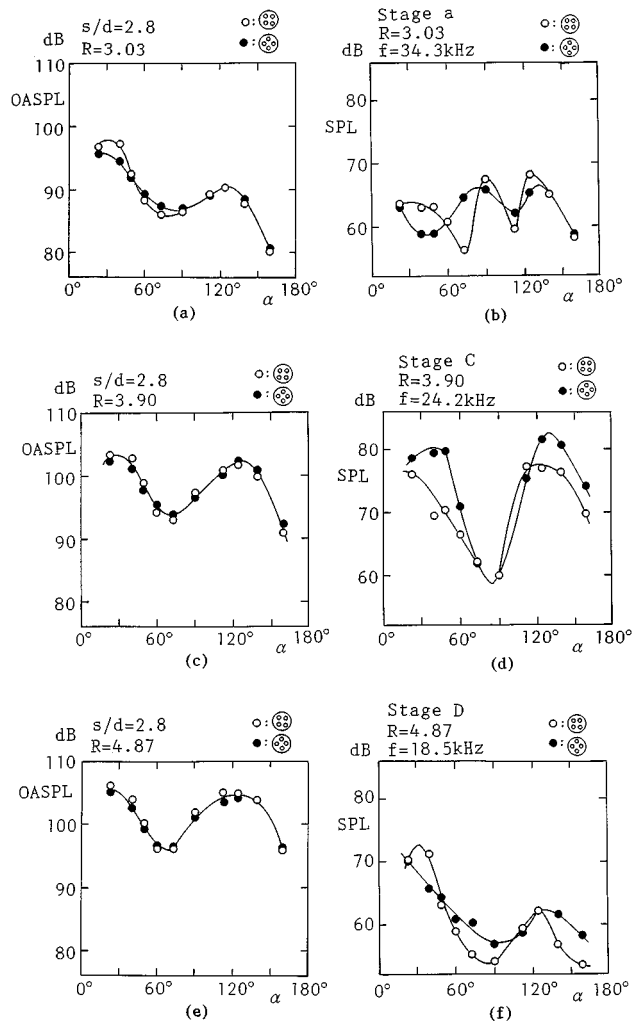


FIG. 13. Directionality of sound radiated from multijets; (a) OASPL at pressure ratio $R=3.03$, (b) SPL for discrete frequency at pressure ratio $R=3.03$, (c) OASPL at pressure ratio $R=3.90$, (d) SPL for discrete frequency at pressure ratio $R=3.90$, (e) OASPL at pressure ratio $R=4.87$, and (f) SPL for discrete frequency at pressure ratio $R=4.87$.

obtained for the multiple jets with a square configuration cannot be predicted from the single jet results.

In the near sound field, the maximum sound-pressure level is reached at the pressure ratio R of about 5 (mode D) for the twin jet and multijet with square configuration. The differences between the maximum sound-pressure levels of these multiple jets and that of the single jet are about 20 and 10 dB, respectively.

But, the OASPL in the far-sound field for a twin jet and the multijet with square configuration differs from that of single jet by about 3 and 6 dB, respectively, at the same pressure ratios. This result shows that these OASPL are proportional to the cross-sectional area of the nozzles.

The directionalities of the OASPL measured at two inclination angles of a line through the center line of the twin jet and of the diagonal line of the multijet with square configuration to the horizontal plane (microphone traversing plane) are very similar to each other at each pressure ratio.

ACKNOWLEDGMENTS

The authors received support from the Ministry of Education and Culture of Japan (Grant No. C-07650202).

- Davies, M. G., and Oldfield, D. E. S. (1962). "Tones From a Choked Axisymmetric Jet. II. The Self Excited Loop and Mode of Oscillation," *Acustica* **12**, 267–277.
- Hammit, A. G. (1961). "The Oscillation and Noise of an Overpressure Sonic Jet," *J. Aerospace Sci.* **28**, 673–680.
- Merle, M. (1956). "Sur la fréquence des ondes sonores émises par un jet d'air à grande vitesse," *C. R. Acad. Sci. (Paris)* **243**, 490–493.
- Powell, A. (1953a). "On the Mechanism of Choked Jet Noise," *Proc. R. Soc. London, Ser. B* **66**, 1039–1056.
- Powell, A. (1953b). "On the Noise Emanating from a Two-Dimensional Jet above the Critical Pressure," *Aeronaut. Q.* **4**, 103–122.
- Powell, A., and Kozu, T. (1991). "Observations of screech of jets emerging from the 90° corner between plane reflectors, pertinent to valve noise," *J. Acoust. Soc. Am.* **90**, 2245 (A).
- Powell, A., Umeda, Y., and Ishii, R. (1990). "The Screech of Round Choked Jets, Revisited," AIAA paper 90-3980 (unpublished).
- Powell, A., and Ishii, R. (1992). "Observation of the oscillation modes of circular choked jets," *J. Acoust. Soc. Am.* **92**, 2823–2836.
- Seiner, J. M., Manning, J. M., and Ponton, M. K. (1988). "Dynamic Pressure Loads Associated with Twin Supersonic Plume Resonance," *AIAA J.* **26**, 954–960.
- Shaw, L. (1990). "Twin-Jet Screech Suppression," *J. Aircraft* **27**, 708–715.
- Suda, H., Manning, T. A., and Kaji, S. (1993). "Transition of Oscillation Modes of Rectangular Supersonic Jet in Screech," AIAA paper 93-4323 (unpublished).
- Umeda, Y., and Ishii, R. (1994). "Oscillation modes of underexpanded jets issuing from square and equilateral triangular nozzles," *J. Acoust. Soc. Am.* **95**, 1853–1857.
- Westley, R., and Wooley, J. H. (1975). "The near field sound pressures of a choked jet when operating in the spinning mode," AIAA 2nd Aero-Acoustics Conference, AIAA Paper 75-479 (unpublished).
- Wlezien, R. W. (1989). "Nozzle Geometry Effects on Supersonic Jet Interaction," *AIAA J.* **27**, 1361–1367.
- Yu, J. C., and Seiner, J. M. (1983). "Nearfield Observations of Tones Generated from Supersonic Jet Flows," AIAA paper 83-0706 (unpublished).

Frequency dependence of broadband propagation in coastal regions

M. Badiey

Ocean Acoustics Laboratory, Graduate College of Marine Studies, University of Delaware, Newark, Delaware 19716-3501

J. Simmen

Ocean Acoustics Program, Office of Naval Research, ONR 3210A, Arlington, Virginia 22217-5660

S. Forsythe

Naval Undersea Warfare Center, Underwater Sound Reference Detachment, Orlando, Florida 32856-8337

(Received 24 December 1996; accepted for publication 25 January 1997)

To study the frequency-dependent spatial and temporal variabilities of sound propagation in coastal regions two experiments were conducted by transmitting sound impulses in the form of M sequences centered from 0.6 to 22 kHz. The site of the first experiment was the Atlantic Generating Station (AGS) where the source–receiver range was 214 m. The site of the second experiment was Delaware Bay where the range was approximately 760 m. The mean water depth was 14–15 m in both locations. Oceanographic data and acoustic data were collected simultaneously in both experiments. It was found that the temporal coherence of the propagated broadband signal changes significantly with pulse center frequency, as well as varying with geographic location and time. Trends of increasing signal decorrelation (between consecutive pulses) with increasing center frequency, and increasing signal decorrelation for paths with increasing number of interface interactions, were observed. It was found that for lower center frequencies there is hardly any decorrelation in signal over several hours, while for signals with center frequencies only a few kHz higher there was substantial decorrelation over times as short as 10 to 20 min. No significant spatial decorrelation of the signal was observed over the hydrophone spacing of a few meters for these experiments. © 1997 Acoustical Society of America. [S0001-4966(97)06305-4]

PACS numbers: 43.30.Bp [JHM]

INTRODUCTION

As the focus in underwater acoustics is shifting towards shallow-water environments, and in particular to coastal regions, questions arise regarding the spatial and temporal coherence of acoustic waves as they propagate in such regions. The source and receiver relative positions, the spectral character of the transmitted waveform, and the temporal and spatial variability of the environment are all important parameters that affect the acoustic coherence. Sound propagation in shallow-water regions is complicated by multiple interactions of the acoustic waves with the sea bottom and sea surface, in addition to the usual interaction with the water column. The degree to which these different interactions cause degradation in the coherence of the pressure field at some distance is not well understood and depends on the scales of the variability in the water column, sea surface, and sea bottom, as well as the frequency of the acoustic signal. Numerical modeling has indicated the importance and interplay of sea surface and bottom roughness alone¹ or of volume fluctuations like internal waves.² Clearly, the ratio between the acoustic wavelength and the dominant scale of spatial variabilities in the shallow-water environment is an important determinant in spatial coherence of the acoustic field; likewise, the temporal variability of the relevant spatial scales in the water column and sea surface controls the temporal coherence of the acoustic field. Spatial and temporal coherence,

in turn, set the performance limitations of sonar apertures, real or synthetic, or determine the applicability of techniques based on phase conjugation or reciprocity.

Ocean environment monitoring techniques that exploit the use of acoustics (e.g., acoustical tomography³) have been practiced in deep water but may work equally as well in shallow water. The success of acoustical tomographic methods in shallow-water environments will be limited by our ability to separate out acoustic fluctuations due to the ocean volume from those due to the sea surface and sea bottom/sub-bottom. Recent studies have begun to examine the potential of acoustical inverse methods in shallow-water coastal regions,^{4–6} and there have been several efforts to correlate the fluctuations in the acoustic signal (e.g., travel time⁷) with temporal and spatial variabilities in the environment. The optimal choice of a signal frequency or waveform to probe shallow-water environments is not obvious and will depend on the goals of the tomographic effort. To address this latter issue, we have initiated a study of the dependence of acoustic fluctuations on signal center frequency for pulses propagated in shallow water. Our immediate objective is to determine the dependence of coherence for a propagated broadband signal on center frequency, ray path, and ocean environment. In this paper we present the first step in a continuing series of such studies in shallow water.

Section I provides a brief description of the initial at-sea experiments, with emphasis on the implemented signal pro-

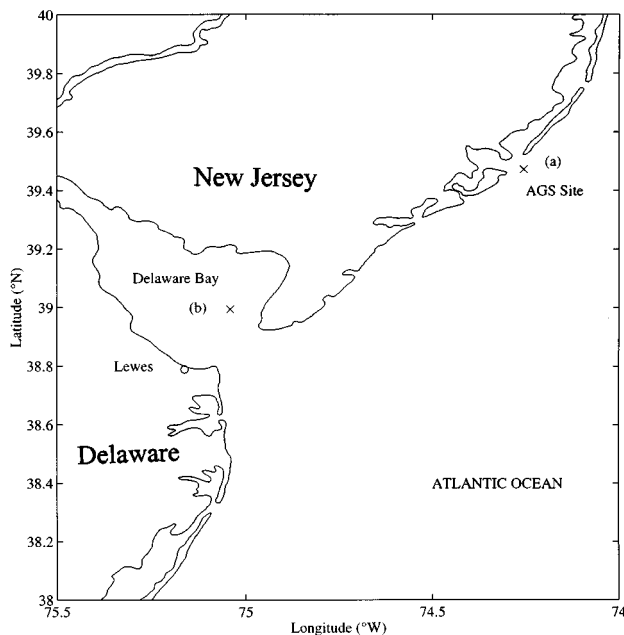


FIG. 1. Map of New Jersey and Delaware coastlines with geographic location of experiments labeled (a) Atlantic Generating Station (AGS) and (b) Delaware Bay.

cessing. Sections II and III each contain for the first and second experiments, respectively, an overview of the acoustic and oceanographic measurements, plus the highlights of the analysis and discussion. Finally in Sec. IV a summary of the results and some thoughts about a future experiment are presented.

I. SIGNAL PROCESSING

Underwater acoustic experiments were conducted in two different shallow-water regions to investigate the frequency dependence of transmitted signal behavior. The first experiment was performed in May 1994 at the Atlantic Generating Station (AGS) site offshore New Jersey in waters having a mean depth of approximately 14 m. The second experiment was conducted in June 1995 inside the Delaware Bay in waters having a mean water depth of approximately 15 m. Both sites have been locations for prior acoustical,⁸ oceanographic,⁹ and geological studies.¹⁰ Figure 1 shows the geographic locations and Fig. 2 presents a general schematic layout of the experiments. In both cases, the source and receiver were mounted on tripods placed on the seafloor. Respective source-receiver ranges for the two experiments were 214 m and approximately 760 m. Series of M sequences were transmitted at different center frequencies to allow an easy examination of pulse propagation behavior as a function of frequency. Oceanographic parameters of temperature and salinity were measured by thermistors and CTDs (Conductivity, Temperature, and Depth).

Extracting the dependence of received time series on frequency requires that a narrow-band signal be used for frequency resolution. In addition, narrow-band signals offer a high signal-to-noise ratio for a given transmission duration. Conversely, localization (isolation of multipaths) in time re-

quires wideband signals. As a result, a variety of pulses that are simultaneously time- and bandlimited were chosen.

For the first (AGS site) experiment, Gaussian-modulated sinusoids described the pulses. The defining equation of the transmitted signal can be expressed by

$$g(t) \propto e^{-k(t-w/2)^2} \cos \omega(t-w/2), \quad (1)$$

where k specifies the time- and bandwidth of the pulse, w is the total pulse width, and $\omega = 2\pi f_c$, where f_c is the center frequency. A pulse width of 1.0 ms (at 0.5 amplitude of the Gaussian envelope) and a bandwidth of 1 kHz were chosen to give good simultaneous frequency and time resolution. The center frequencies ranged from 0.6 through 7 kHz. For the second (Delaware Bay site) experiment, broadband pulses with variable bandwidths of 0–4 kHz, 0–8 kHz, 5–13 kHz, and 9–25 kHz were used. These bands were chosen to cover the entire low-to-midfrequency spectrum with overlapping coverage. The broadband pulses were filtered to divide the original 4 bands into Gaussian-shaped sub-bands with a fixed timewidth and center frequencies ranging from 1 to 22 kHz. The bandwidths of the lowest frequencies were narrower than those of the higher frequencies to insure that behaviors such as bottom interaction would not be averaged across too wide a band. For example, for 1–2 kHz center frequency the bandwidth was 1 kHz, for 3–6 kHz it was 2 kHz, for 7–11 kHz it was 3 kHz, and for 12–22 kHz the bandwidth was 4 kHz.

In both experiments, the pulses were further modulated by convolution with length 511 M sequences^{11,12} to insure that maximum signal power per unit time would be delivered into the water while simultaneously retaining the spatial discrimination of the original pulses. The resulting signals were transmitted continuously for periods ranging from 5 to 20 s. To increase the effective signal-to-noise ratio of the arriving signal components, coherent time averaging was employed.

After recovering the original pulses by matched filtering, the time series data were reduced to complex baseband signals and down-sampled to save disk storage. This process removes the high-frequency carrier and makes a comparison of pulse shapes easier across frequency. All graphic representations of the signal time series in this paper are the amplitudes of the complex time series.

II. AGS EXPERIMENT

The experiment at the AGS site was the first and simplest of a planned sequence of experiments to investigate the frequency dependence of signal coherence in very shallow water. The duration of the acoustic experiment was only four hours, and the maximum pulse center frequency was only 7 kHz. The site was selected because of some familiarity with the acoustical properties there⁸ and it is thought to be representative of eastern seaboard, continental-shelf, shallow-water environments. For the limited dataset of the AGS experiment we highlight a few immediate, yet significant, observations.

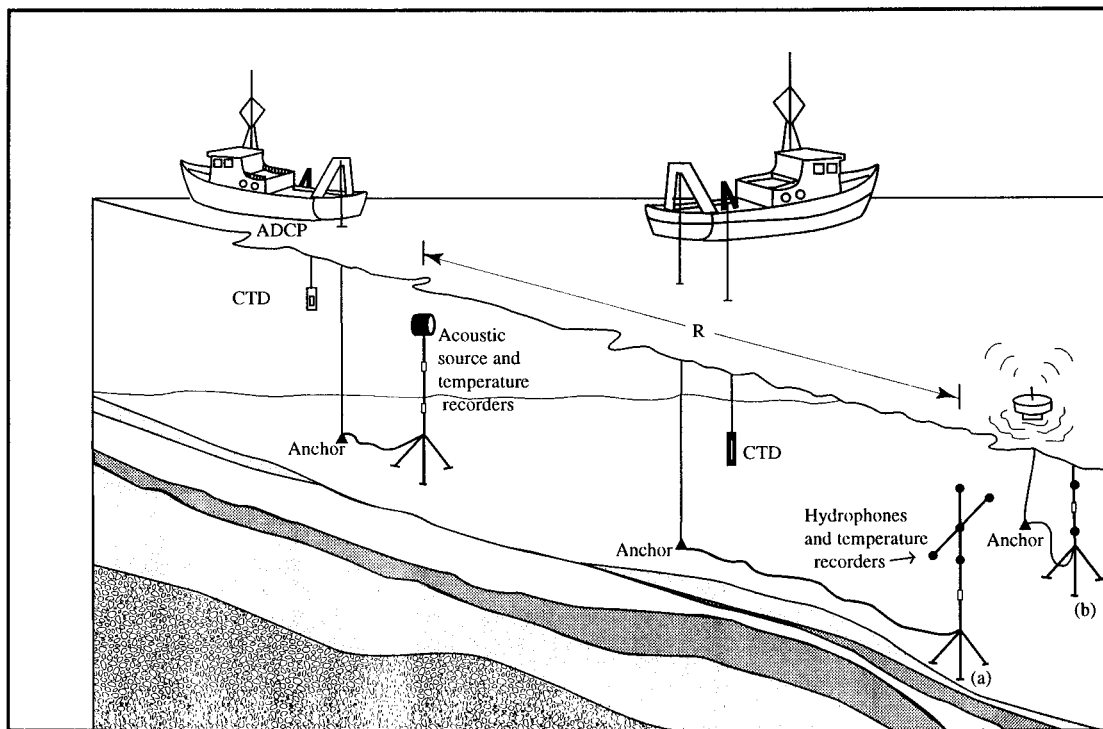


FIG. 2. Experimental configuration: (a) Atlantic Generating Station site (range, $R = 214$ m) and (b) Delaware Bay site ($R \approx 760$ m).

A. Measurements

The acoustic receiver consisted of five hydrophones assembled in the form of a cross (see Fig. 2). The horizontal and vertical spacing between adjacent hydrophones was one meter, and the center hydrophone was located 2 m from the seafloor. The acoustic source was a J11-3 transducer,¹⁴ also fixed 2 m above the seafloor, and the source and receiver separation was 214 m. The source level was 167 dB *re*: 1 μ Pa at 1 m over a frequency range of 0.6–7 kHz.

The experiment was carried out on the afternoon of May 29, 1994. Ten transmissions of a specified set of pulses (8 center frequencies from 0.6 to 7 kHz) were made and the results were recorded¹³ on the five receiving hydrophones. The interval between transmissions was typically about 20 min. During this experiment, synchronization of the five channels of data recorded by the receiver to the start of the transmission was unsuccessful. This resulted in uncertainty in the absolute times of signal component arrivals, relative to start of transmission. Subsequent analysis for this experiment could not exploit absolute arrival times; however, other characteristics (e.g., correlation versus lag time) of the acoustic propagation independent of the absolute time could still be determined.

Sound speed (temperature and salinity) and water column height were measured for the duration of the acoustical experiment. CTDs, one near the source and the other near the receiver, were taken hourly. Tide data for the area indicated a change in sea level of about 1 m over 5 h and since the experiment was conducted in 14 m of water, it was deemed important to tide-correct the depths recorded by the CTD instruments. Figure 3 displays the sound speed as observed

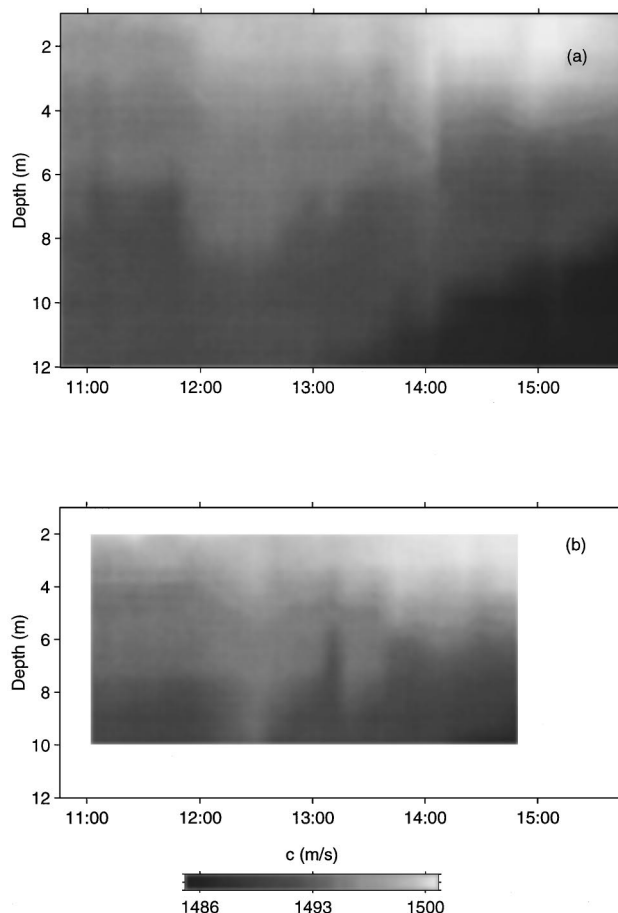


FIG. 3. Contour plots of sound speed versus time and depth at AGS site: (a) near source and (b) near receiver.

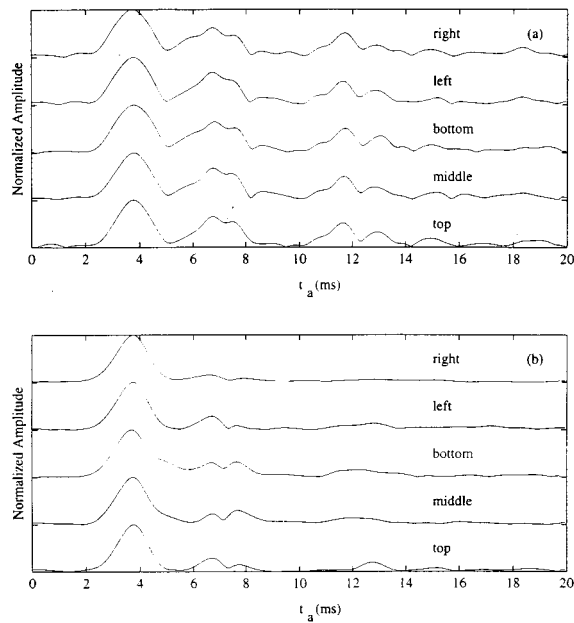


FIG. 4. Signal amplitude versus arrival time t_a for all five hydrophones: (a) $f_c = 1$ kHz and (b) $f_c = 7$ kHz.

by the source and receiver stations showing significant variability in the (downward refracting) sound-speed profile.

B. Analysis and observations

Figure 4 displays, for the first pulse transmission and two different center frequencies ($f_c = 1$ and 7 kHz), the received signal amplitude versus arrival time (t_a) at all five hydrophones. The multiple peaks of each reception are due to time differences in arrivals corresponding to different ray paths. For the higher frequency the late-arriving (boundary interacting) peaks are more attenuated. For all pulses formed in this experiment (i.e., for center frequencies up to 7 kHz) the observed constancy of the receptions from hydrophone to hydrophone is exceptional. In fact, there is only a small change in spatial coherence over 2 m separation as the center frequency is varied from 0.6 to 7 kHz. Note that this high spatial coherence only applies to the short distances between hydrophones realized in this experiment, and one would expect a further spatial decorrelation as the distance between hydrophones or the source–receiver range increases.

Figure 5 displays the signal amplitude versus arrival time for all ten pulse transmissions at each of two center frequencies, 1 and 7 kHz, for the middle hydrophone. For the low center frequency of 1 kHz [Fig. 5(a)] the received signal is very stable, showing minor variation as a function of pulse number. At the higher 7 kHz [Fig. 5(b)] center frequency, instead, there is noticeable variability from pulse to pulse. There does, however, appear to be elements of periodicity for the 7-kHz case. For example, pulse numbers 2 and 7, and pulse numbers 5 and 10, bear a strong resemblance to each other. This apparent periodicity does not correlate to tide and must be related to other oceanographic phenomena.

Figure 6 shows the correlation versus lag time for three consecutive center frequencies 1, 2, and 3 kHz. By aligning

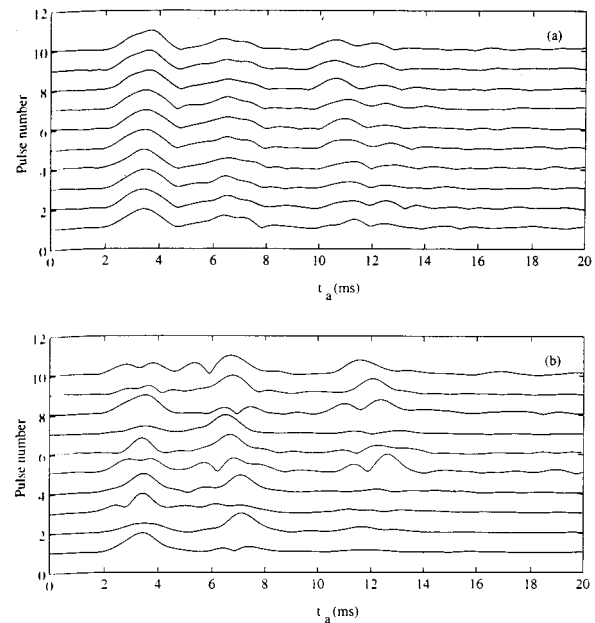


FIG. 5. Signal amplitude versus arrival time t_a for all pulses and the middle hydrophone: (a) $f_c = 1$ kHz and (b) $f_c = 7$ kHz. The interval between consecutive pulses was typically about 20 min.

the first arrivals of pulse numbers i and j to produce shifted complex time series, $S_i(t)$ and $S_j(t)$, the correlation matrix C_{ij} is calculated from

$$C_{ij} = \left\| \sum_k S_i(t_k) S_j^*(t_k) \right\| / \left(\sum_k \|S_i(t_k)\|^2 \right)^{1/2} \left(\sum_k \|S_j(t_k)\|^2 \right)^{1/2}, \quad (2)$$

where the t_k 's are the discretized (sampled) times of the complex time series. In practice, small offsets in alignment are made until the correlation is a maximum, thus allowing

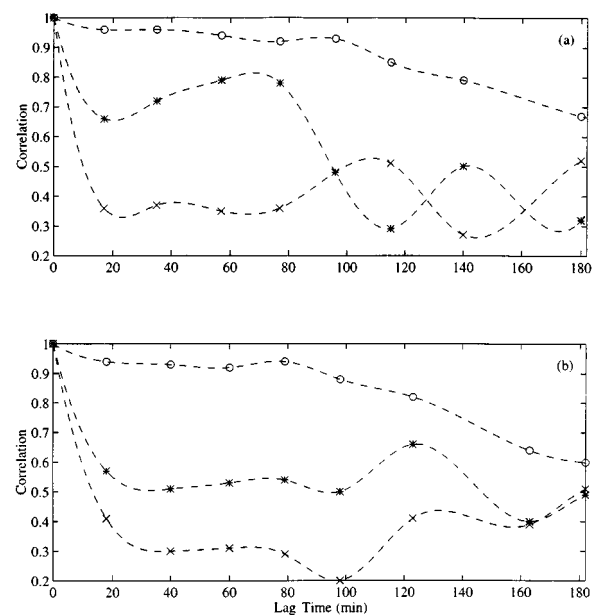


FIG. 6. Correlation versus lag time for three separate center frequencies: $\circ = 1$ kHz, $*$ = 2 kHz, and $x = 3$ kHz; (a) lag relative to first pulse and (b) lag relative to second pulse.

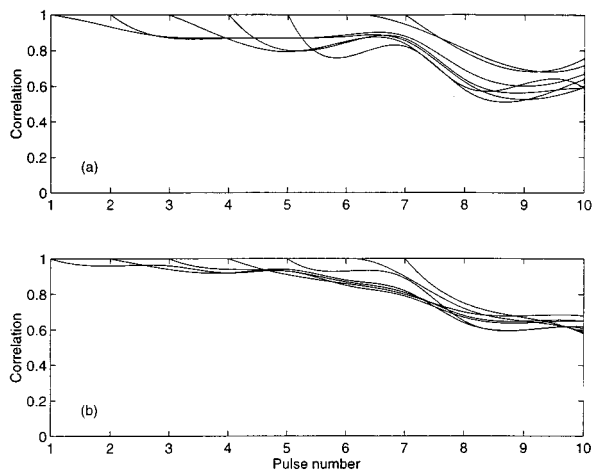


FIG. 7. Correlation of each pulse with subsequent pulses (for pulse numbers 1–7). (a) $f_c = 0.6$ kHz and (b) $f_c = 1.0$ kHz.

for small uncertainties in the initial alignment process. In Fig. 6(a) and (b), i is fixed to be 1 and 2, respectively, j varies from i to $i + 7$, and the correlation in each case is plotted against lag time, which is just the geotime separating pulses i and j . Note that for 1 kHz the signal decorrelates slowly in lag time; on the other hand, for 2 kHz there is noticeable decorrelation at about 20 min lag time, and for 3 kHz there is even more immediate decorrelation. Further analysis shows that for center frequencies above 3 kHz this immediate decorrelation is persistent. Between 1- and 3-kHz center frequencies is where a transition away from temporal “stability” occurs.

It is of interest to look at the decorrelation of each pulse by comparing subsequent pulses. For the lower center frequencies, affected by the larger environmental scales, there is a strong correspondence between the time of significant oceanographic change and the geotime when the strong decorrelation occurs. Figure 7 shows that between pulse numbers 7 and 8 ($T_g \approx 13:30$ h) at $f_c = 0.6$ and 1 kHz there is an obvious drop in correlation: this corresponds to drastic changes in the vertical sound-speed structure commencing at the same geotime (see Fig. 3).

III. DELAWARE BAY EXPERIMENT

The experiment in Delaware Bay was an extension of the first experiment; it spanned a longer time period (i.e., 12 h) and it allowed pulse center frequencies to vary from 1 to 22 kHz. Furthermore, because of successful synchronization, absolute times of signal component arrivals could be used in the analysis. Since only one receiving hydrophone recorded continuously, however, spatial coherence could not be examined in this experiment. A more extensive analysis than that of the first experiment was conducted for the Delaware Bay dataset.

A. Measurements

The receiver array in this experiment was tethered to a surface buoy located approximately 760 m from the source (Fig. 2), and the received acoustic signals were transmitted from the surface buoy to the source ship via VHF radio in

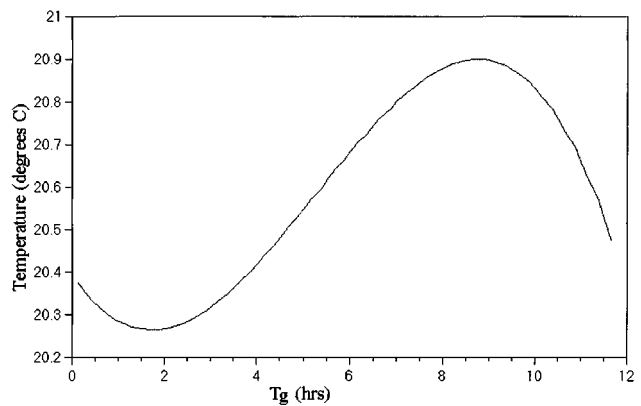


FIG. 8. Spatially averaged temperature versus geotime T_g .

163.5–174 MHz band. Only one of the receiver hydrophones functioned successfully for the duration of the experiment. The acoustic source was an F56 transducer with a maximum power¹⁵ of 160 dB *re*: 1 μ Pa at 1 m.¹⁴ This source was tethered to the moored ship (R/V CAPE HENLOPEN).

This experiment began on 29 June 1995 and lasted from approximately 6 p.m. to 6 a.m. the following morning. 140 transmissions of a specified set of pulses, having center frequencies from 1 to 22 kHz, were made. During this experiment, it was possible to synchronize all hydrophone channels with the transmitted data, resulting in an absolute time base upon which the time-of-arrival information for all channels was based.

In addition to temperature measurements taken at both the source and receiver, other oceanographic measurements were made. Several CTD casts per hour at the R/V CAPE HENLOPEN reported temperature and salinity measurements while concurrent ADCP (Acoustic Doppler Current Profiler) measurements on the same ship recorded current profiles. Also, wind speed and ship position, using a differential global positioning system (DGPS), were measured.

B. Analysis and observations

Figures 8 and 9 summarize the most relevant oceanographic measurements for this experiment. In both figures sinusoidal variation due to tide is evident. Figure 8 shows the temperature averaged over source and receiver records versus geotime, T_g , for the experiment’s duration. Figure 9(a) shows tide height versus geotime extrapolated to the experiment site from measurements at a nearby location. Figure 9(b) shows the corresponding average current, based on ADCP measurements at three different depths near the source. Note the 90° phase shift from the tidal height curve. The origin of the T_g axis in these figures corresponds to 6 p.m. local time on 29 June 1995.

During the acoustic experiment, strong winds forced unexpected and significant changes in the ship’s position, which in turn dragged the source tripod to new positions. Additionally, the source tripod was immediately tipped, putting the source on the seafloor. This source–receiver relative motion needed to be accounted for in the analysis of the acoustic fluctuations. Figure 10 shows, for a center frequency of 2 kHz, the received signal amplitude versus arrival

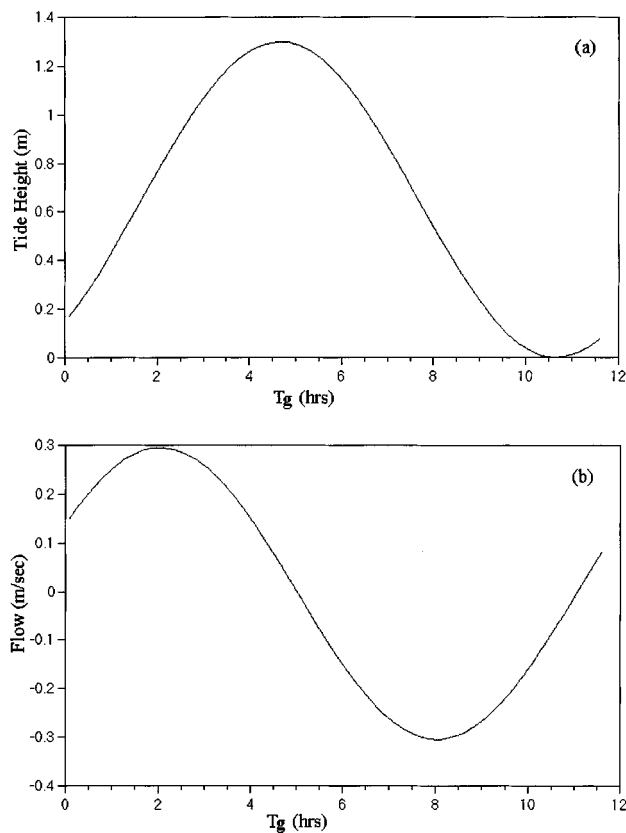


FIG. 9. Tide height (a) and current speed (b) versus geotime along the source-receiver line (+ means the flow is toward the source).

time for a sequence of pulses (sequence of geotimes). At $T_g = 5.5$ h a drastic change in the peak arrival times is obvious; these changes, which calm within a few hours, correspond to this source motion.

The changes (as geotime varies) in travel time for the first arrivals of the pulses shown in Fig. 10 are due to variations of the ocean sound speed and source-receiver range, and to a much lesser extent, changes in the ocean currents. Because the source tripod was cabled to the ship, and the cable could have slack, there was no direct way to measure the source motion. Since the first arrival is along the direct

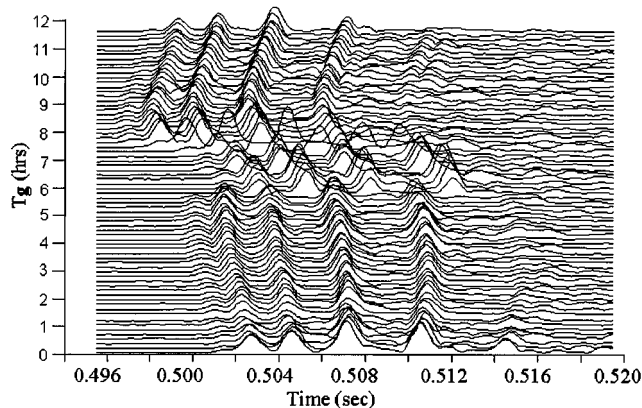


FIG. 10. Received signal amplitude versus arrival time for 2 kHz center frequency. These are the time series prior to correcting for source-receiver motion.

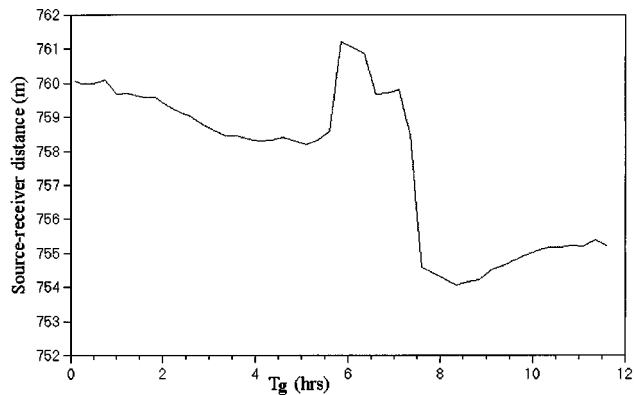


FIG. 11. Estimated source-receiver separation as a function of geotime.

path, its travel time is independent of ocean depth. Since the effects due to current changes are high order, accounting for the change in sound speed over geotime given by Fig. 8, the source-receiver distance as a function of geotime can be deduced from the first arrivals. In Fig. 11, the inferred source-receiver separation (based on the 9-kHz center frequency data) is shown. Several meter changes in separation correlate well with recorded changes in the ship's position.

Simple ray-based calculations allow us to interpret the received data from a geometrical point of view. Figure 12 diagrams the paths for rays having zero boundary interactions (direct path) to only a few. Because of the geometry, the ray paths arrive in interfering groups. Figure 12(a) shows, for example, the three ray paths that comprise the first interfering group, including the direct path (DP), single surface bounce (S), and surface-bottom bounce (SB). Subsequent interfering groups arrive in pairs, corresponding to ray paths having multiple surface and bottom interactions (SBS, SBSB,...). For this experimental geometry, the arrival times of the ray paths, arranged by group, are shown in Table I. Notice that in each group the arrival times differ by a fraction of a millisecond (i.e., less than a pulse width) and are therefore interfering.

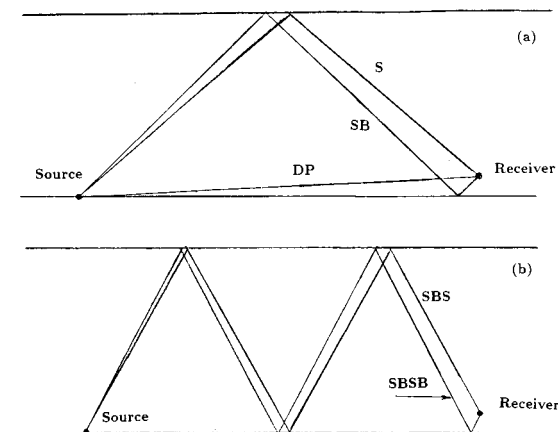


FIG. 12. Ray path diagram showing paths with increasing boundary interaction. (a) First group of arrivals comprised of direct path (DP), single surface bounce (S), and surface-bottom bounce (SB). (b) Second group of arrivals comprised of surface-bottom-surface bounce (SBS) and surface-bottom-surface-bottom bounce (SBSB).

TABLE I. Arrival times for the peaks 1–7 for a typical pulse. D means direct path, S means surface bounce, and B means bottom bounce.

| Peak | Path | Time (ms) |
|------|----------------|-----------|
| 1 | D | 496.64 |
| 1 | S | 497.13 |
| 1 | SB | 497.15 |
| 2 | SBS | 498.32 |
| 2 | SBSB | 498.36 |
| 3 | SBSBS | 500.31 |
| 3 | SBSBSB | 500.36 |
| 4 | SBSBSBS | 503.09 |
| 4 | SBSBSBSB | 503.15 |
| 5 | SBSBSBSBS | 506.64 |
| 5 | SBSBSBSBSB | 506.71 |
| 6 | SBSBSBSBSBS | 510.94 |
| 6 | SBSBSBSBSBSB | 511.03 |
| 7 | SBSBSBSBSBSBS | 515.98 |
| 7 | SBSBSBSBSBSBSB | 516.09 |

Figure 13 shows, for a center frequency of 2 kHz and many different geotimes, the received signal amplitude versus arrival time, after aligning the first arrivals based on a correction for the source–receiver relative motion and changes in the average sound speed. The solid curves in Fig. 13 show the calculated times for the first several groups of arrivals as a function of geotime, based primarily on the changing geometry of ray paths due to tide height [Fig. 9(a)] and secondarily on the advection of sound due to the ocean current [Fig. 9(b)]. The first peak is actually comprised of three interfering arrivals, and subsequent peaks are comprised of pairs of interfering arrivals. Note how closely as a function of geotime the calculated arrival time tracks the

measured arrival time for the first several arriving peaks. The strongest arriving peaks (2–5) are those immediately following the first peak (1), followed by peaks (6,7,...) with noticeably reduced amplitudes, presumably due to the extended interaction with, and attenuation from, the sea surface and bottom.

In Fig. 14 the predicted amplitude versus time is shown using the parabolic equation (PE) to simulate the propagation of a pulse with a 2-kHz center frequency in a simple waveguide. In order to understand the deterministic character of the arrival pattern for the lower center frequencies, this PE calculation was performed in the absence of surface or bottom roughness. The seafloor was modeled as a fluid with a constant attenuation of 0.2 dB per wavelength. Although this calculation is unable to capture the observed stochastic component of the arrivals, it does confirm some important features of the experimental data for the lower center frequencies. Figure 14(a) shows the predicted and measured amplitude versus time for a representative geotime. It shows good agreement in the relative amplitudes of peak arrivals between the prediction and the actual data. The PE calculation confirms that the later arrivals in Fig. 13 (i.e., 6,7,...) are noticeably weaker in amplitude than the earlier ones because of energy loss due to multiple bottom interactions at higher grazing angle. Figure 14(b) shows for many different geotimes, the predicted signal amplitude versus time, and is presented for comparison to the field data of Fig. 13. It shows that except for arriving peak 6 (see Fig. 13), where water depth significantly influences amplitude, the primary effect on the received signal of the changing tide is to change the peak time of arrival.

Figure 15(a)–(d) shows the received signal amplitude versus time for center frequencies of 4, 9, 15, and 20 kHz, respectively. As this progression of figures shows, the later

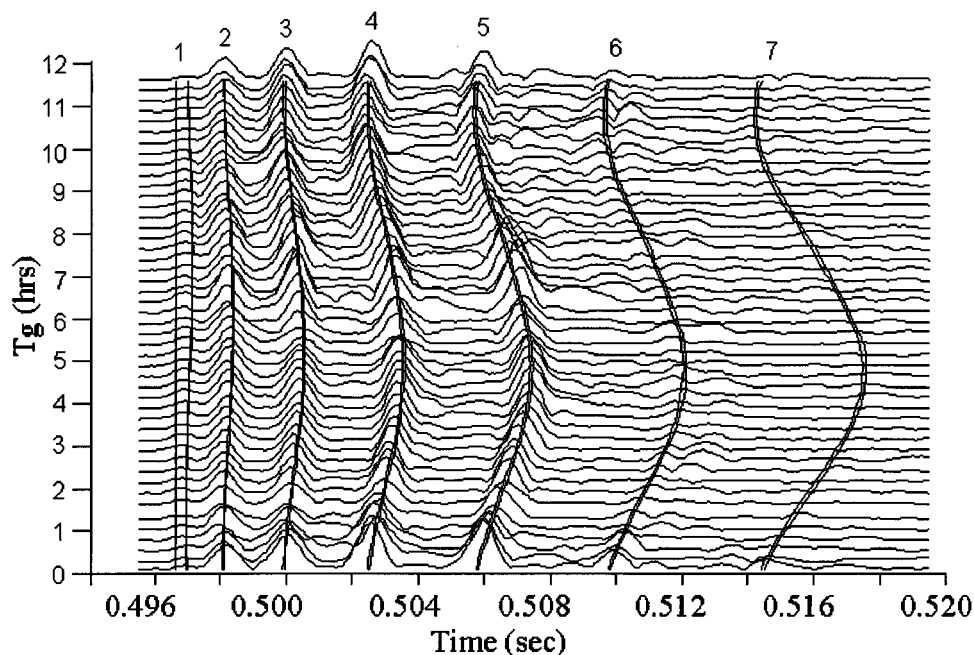


FIG. 13. Measured signal amplitude versus time, after correction for source–receiver separation ($f_c = 2$ kHz). The solid lines are predicted arrival times versus geotime for the first several ray paths corresponding to peaks 1–7.

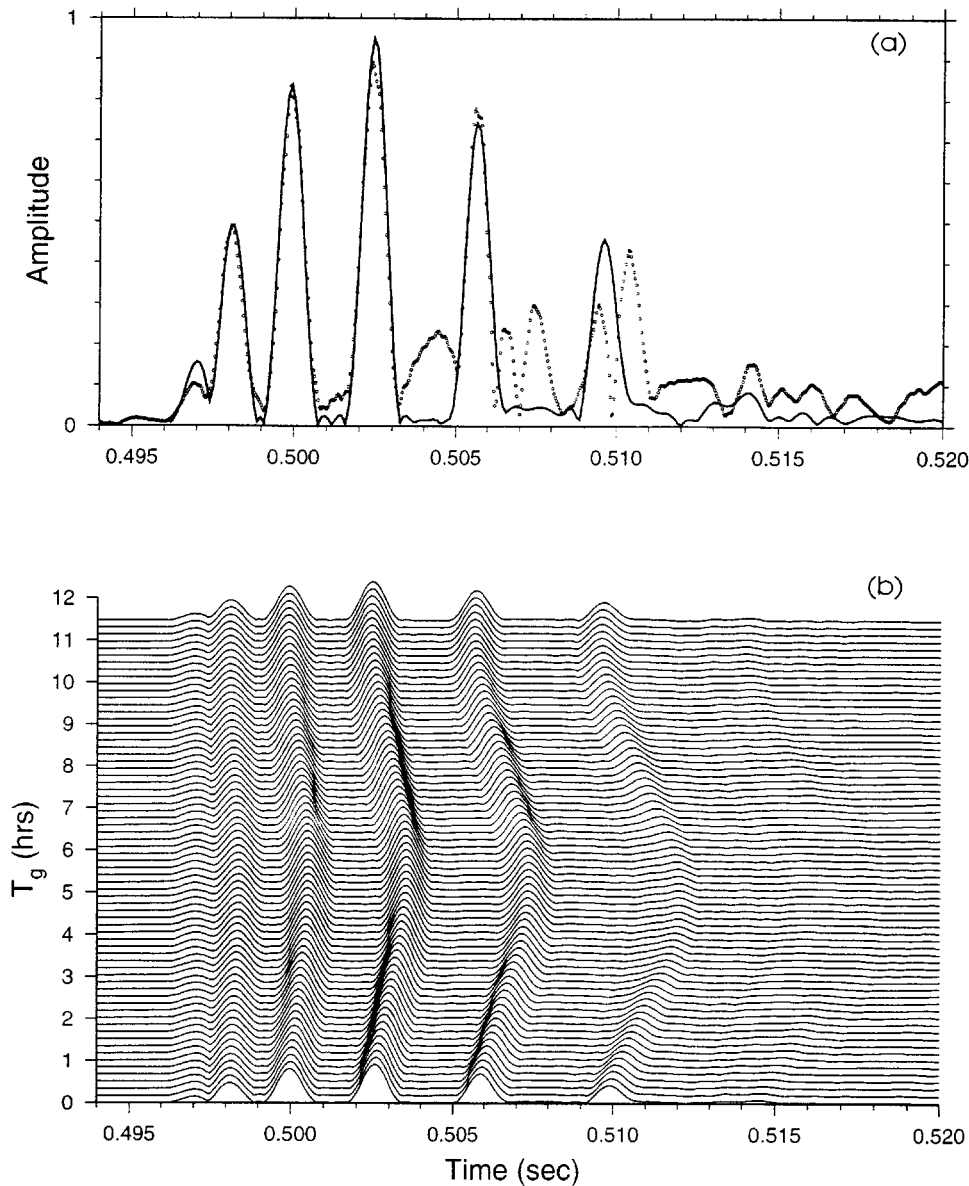


FIG. 14. (a) PE-predicted (solid) and measured (dotted) signal amplitude versus time (for $f_c = 2$ kHz) for a representative T_g . (b) PE-predicted amplitude versus time as a function of geotime.

arrivals are more attenuated for higher frequencies. For example, in Fig. 15(b), corresponding to 9 kHz, there are two fewer strong arrivals than in Fig. 13, corresponding to 2 kHz. At higher frequencies the loss from the surface and bottom bounces is more pronounced, probably because the realized scales of interface roughness accentuate nonspecular scattering at the shorter wavelengths. The monotonic growth of the first peak with frequency is a consequence of the fact that a different variable bandwidth filtering and normalization was applied to each frequency. In addition to dependence on the center frequency or ray paths, there is also a strong dependence of signal amplitude on geotime, and thus, on ocean environmental variability. Figures 15(c) and (d) show, for example, a second peak that is initially faded or broken up until geotime $T_g \approx 7$ h, at which time it becomes pronounced. It may be that surface roughness changed to a different dominant length scale, changing the magnitude of the scattering or attenuation at that center frequency. Unfortunately, there

was no direct monitoring of ocean surface (wave spectra) during this experiment.

Figure 16 illustrates the frequency dependence of the received signals. Figure 16(a)–(c) shows the received signal amplitude versus arrival time as function of center frequency for three different geotimes, corresponding to the start, middle, and end of the experiment, respectively. This figure suggests an “inverse law”: the number of relatively strong arrivals is approximately inversely proportional to the center frequency. It also shows clearly how the first peak, consisting of three close arrivals, monotonically grows in amplitude with increasing center frequency. Comparisons between Fig. 16(a)–(c) show that the dependence of the second peak on frequency, also changes notably with geotime. In Fig. 16(b) the second peak attenuates strongly for center frequencies beyond 12 kHz, while in Fig. 16(c) it remains pronounced, though broken up some, for all relevant frequencies. Since this peak corresponds to energy having two interactions with

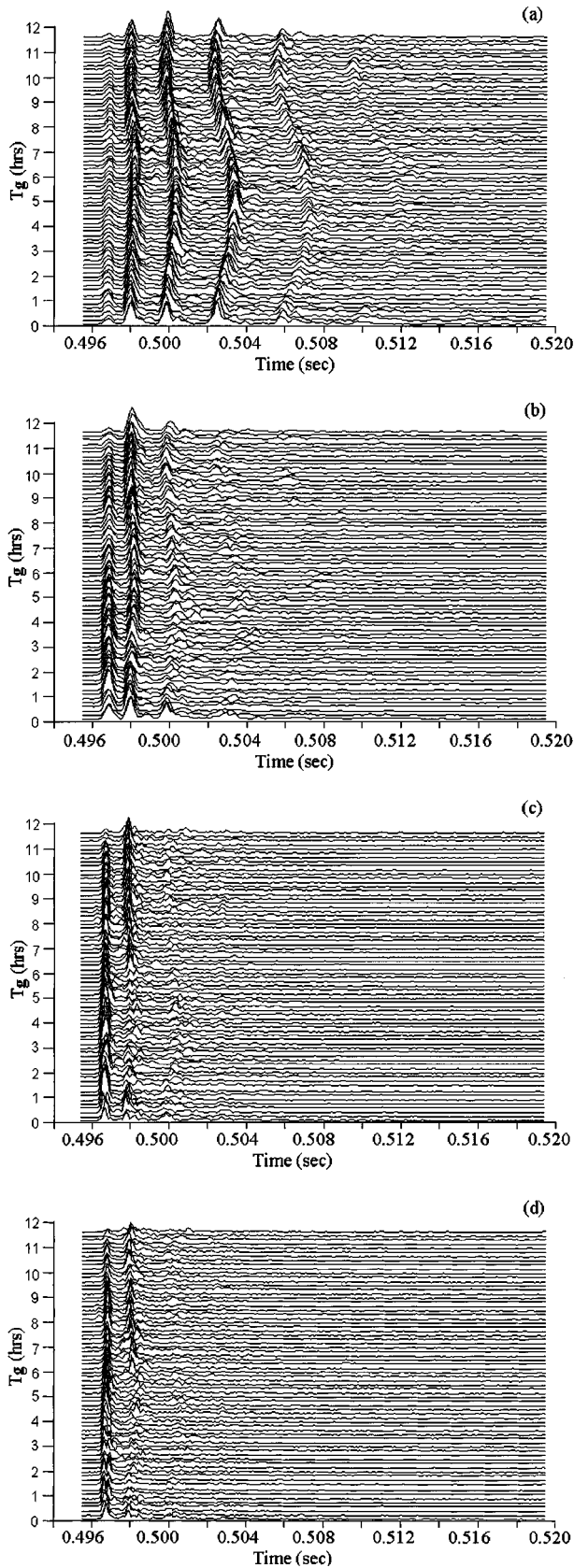


FIG. 15. Measured signal amplitude versus time, after correction for source–receiver separation for (a) $f_c = 4$ kHz, (b) $f_c = 9$ kHz, (c) $f_c = 15$ kHz, and (d) $f_c = 20$ kHz.

the sea surface, it is likely that a change in the roughness in the sea surface is responsible for the differences. For fre-

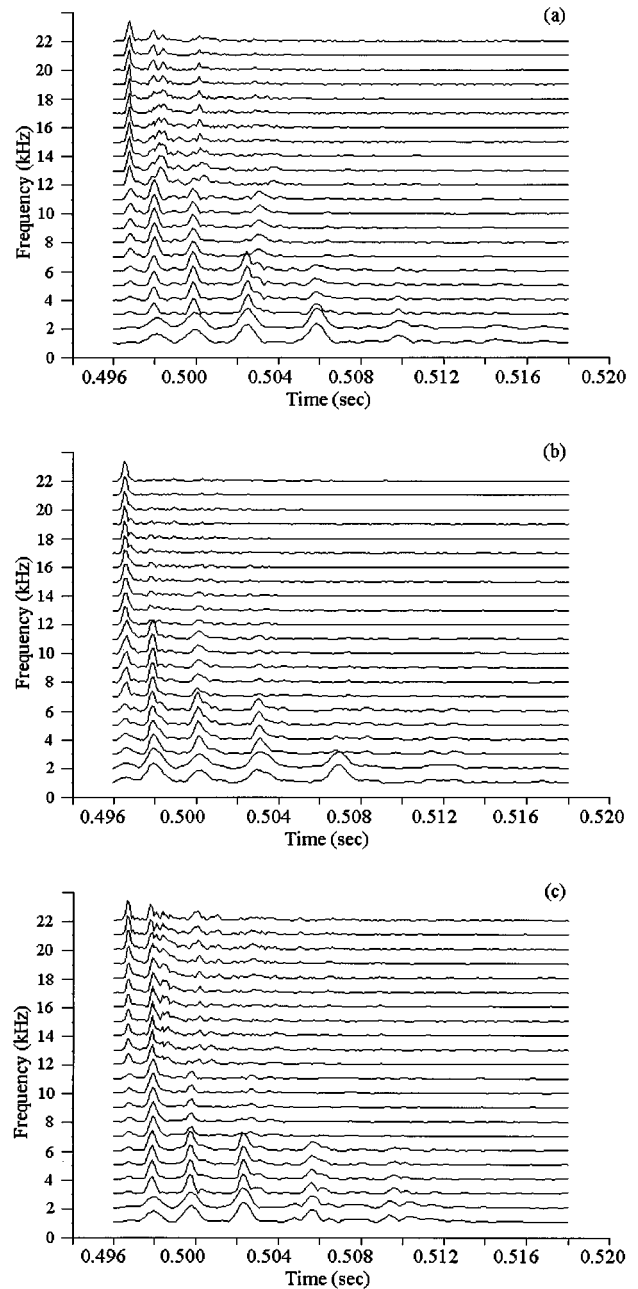


FIG. 16. Time series of received signal amplitude as a function of frequency for a fixed geotime: (a) $T_g \approx 0$ h, (b) $T_g \approx 5$ h, and (c) $T_g \approx 12$ h.

quencies less than 12 kHz the second peak appears to be stable for all geotimes. Also note that for the lower frequencies, peak 6 is visible at the geotimes corresponding to low tide [Fig. 16(a) and (c)] and is lost in the noise at the geotime near high tide [Fig. 16(b)]. Increasing tide height changes the grazing angle which, in turn, increases attenuation, and this is suggested by modeling [see Fig. 14(b)].

One means of quantifying acoustic variability is to compute correlation versus lag time. Figure 17 shows the average correlation versus lag time (for the second peak), at several center frequencies and a fixed ray path [Fig. 17(a)], or for several ray paths at a fixed center frequency [Fig. 17(b)]. The average correlation, $C_{ave}(\tau)$, is computed according to

$$C_{ave}(\tau) = \frac{1}{M} \sum_{|i-j|=k} C_{ij}, \quad (3)$$

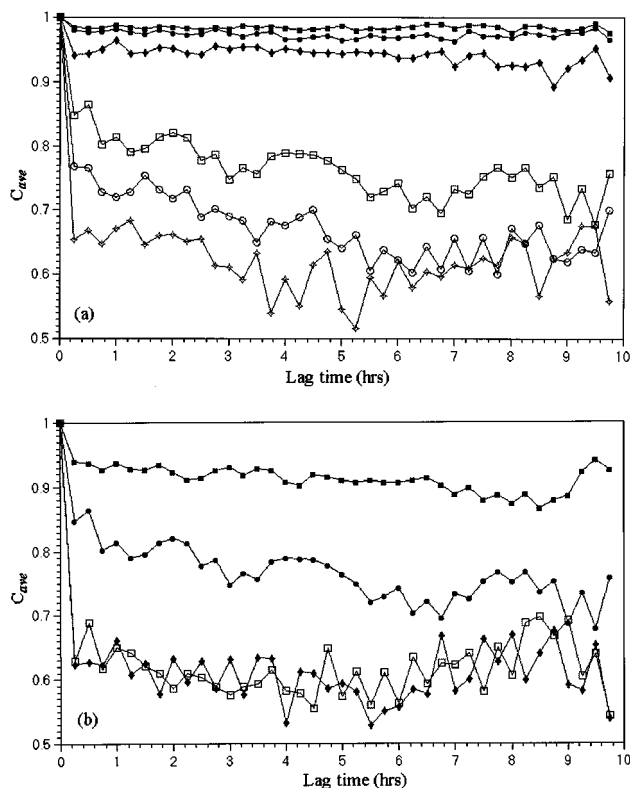


FIG. 17. (a) Average correlation versus lag time for $f_c = 2, 4, 9, 12, 15,$ and 20 kHz: 2 kHz (■), 4 kHz (●), 9 kHz (◆), 12 kHz (□), 15 kHz (○), and 20 kHz (◇). The correlation is for the peak corresponding to the second group of arriving rays (e.g., peak 2 in Fig. 13). (b) Average correlation versus lag time for the first 4 peaks at $f_c = 12$ kHz: peak 1 (■), peak 2 (●), peak 3 (◆), and peak 4 (□).

where M is the number of pulse pairs (i, j) such that $|i - j| = k$, and the lag time τ is just the geotime difference, $|t_i - t_j|$, dependent on k only. Here C_{ij} is calculated from segments of the time series associated with a particular ray arrival. Figure 17(a) confirms the rapid decorrelation at higher frequencies, especially those above 9 kHz. Figure 17(b) shows that the peak formed by the first group of three arrivals remains highly correlated over long lag times, whereas subsequent peaks show faster decorrelation.

IV. CONCLUSIONS

Based on data taken during two pulse propagation experiments, it has been shown that in very shallow water there is a strong frequency dependence of the temporal coherence for signals transmitted over several hundred meters. Typically, above center frequencies of only a few kHz there is substantial decorrelation within a few tens of minutes. There are also strong differences in the temporal stability of the various component arrivals corresponding to different ray paths. Those paths that interact with the seafloor and sea surface are most susceptible to immediate decorrelation for center frequencies of a few kHz and higher. Furthermore, we have shown that all of these effects can strongly depend on the ocean environment at the time and place of the acoustic transmissions. Initial modeling using ray-based and PE approaches has captured the deterministic behavior of the ex-

perimental results. Further modeling efforts will need to account for the stochastic behavior of the ocean, including the bottom and surface roughness.

Future experiments in this sequence to explore the frequency dependence of acoustic fluctuations should require that: (i) the source and receiver positions remain fixed for the duration of the experiment; (ii) there are multiple hydrophones receiving the acoustics transmissions so that all components of spatial coherence can be measured; (iii) the duration of the experiment is several days to allow a reasonable statistical analysis for the collected data and to cover a wider range of geotimes; (iv) surface roughness (wave spectra) be measured so as to be able to correlate changes in coherence with changes in surface roughness; and (v) absolute travel times be measured to allow a more complete analysis. Improvements (iii) and (iv) will permit realistic inputs to propagation models so that a stochastic comparison of model to measurements can be made.

ACKNOWLEDGMENTS

This work was performed partially under the Scientific Officers Research Program (SORP) at ONR while the first two authors worked there and later with partial support of the Sea Grant. The authors wish to thank the crew and the captain of the R/V CAPE HENLOPEN for their help in the experiments.

- ¹D. Rouseff and T. Ewart, "Effect of random sea surface and bottom roughness on propagation in shallow water," *J. Acoust. Soc. Am.* **98**, 3397–3404 (1995).
- ²J. 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).
- ³W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge U.P., New York, 1995).
- ⁴C.-S. Chiu, J. H. Miller, W. W. Denner, and J. F. Lynch, "Forward modeling of the Barent sea tomography vertical line array data and inversion highlights," *Full Field Inversion Methods in Ocean and Seismo-Acoustics* (Kluwer, Dordrecht, 1995), pp. 237–242.
- ⁵H. A. DeFerrari and H. B. Nguyen, "Acoustic reciprocal transmission experiments, Florida Straits," *J. Acoust. Soc. Am.* **79**, 299–315 (1986).
- ⁶H.-H. Essen, F. Schirmer, and S. Sirkes, "Acoustic remote sensing of internal waves in shallow water," *Int. J. Remote Sensing* **4**, 33–47 (1983).
- ⁷J. Lynch, J. Guoliang, R. Pawlowicz, D. Ray, A. Plueddemann, C.-S. Chiu, J. Miller, R. Bourke, A. R. Parsons, and R. Muench, "Acoustic travel-time perturbations due to shallow-water internal waves and internal tides in the Barents Sea Polar Front: Theory and experiment," *J. Acoust. Soc. Am.* **99**, 803–821 (1996).
- ⁸M. Badiey, I. Jaya, and A. H.-D. Cheng, "Shallow water acoustic/geoacoustic experiments at the New Jersey Atlantic Generating Station Site," *J. Acoust. Soc. Am.* **96**, 3593–3604 (1994).
- ⁹S. Glenn and M. Crowley, "Coastal upwelling and its relation to hypoxia in New York bight," *American Geophysical Union, Suppl. to EOS*, S183, April 1995.
- ¹⁰J. Ewing, J. A. Carter, G. H. Sutton, and N. Barstow, "Shallow water sediment properties derived from high frequency shear and interface waves," *J. Geophys. Res.* **92**, 4739–4762 (1992).
- ¹¹An M sequence is a pseudo-random repetitive sequence with repeat period of $N = 2^k - 1$ whose autocorrelation is a delta function.
- ¹²S. W. Golomb, *Shift Register Sequences*, revised ed. (Aegean Park Press, Laguna Hill, CA, 1982).
- ¹³Records of 21.5 s were averaged during the processing.
- ¹⁴L. E. Ivey, "Underwater electroacoustic transducers," Naval Research Laboratory, Underwater Sound Reference Detachment, NRL/PU/5910–940267, 1994.
- ¹⁵The average signal-to-noise ratio for this experiment was calculated to be 25 dB.

All-frequency normal-mode solution of the three-dimensional acoustic scattering from a vertical cylinder in a plane-horizontal waveguide

Gerassimos A. Athanassoulis and Konstadinos A. Belibassakis

*Department of Naval Architecture and Marine Engineering, National Technical University of Athens,
P.O. Box 64070, Zografos, Athens, Greece*

(Received 19 February 1996; revised 16 January 1997; accepted 18 February 1997)

The three-dimensional acoustic scattering from a vertical, impenetrable cylinder in a waveguide is studied. The analytical solution of the problem, for a Dirichlet or a Neumann boundary condition on the scatterer, has been derived recently by Athanassoulis and Prospathopoulos [J. Acoust. Soc. Am. **100**, 206–218 (1996)] in the form of a double-infinite normal-mode series, representing the total acoustic field. In order to extend the applicability of this solution to higher frequencies, the total field is decomposed into the incident and the scattered parts. A series expansion for the scattered field is obtained, and the critical parameter controlling its azimuthal convergence is shown to be the nondimensional wave number ka based on the radius a of the cylinder. The general term of the series starts to decay exponentially immediately after the azimuthal index has exceeded the critical value ka , a fact justifying the introduction of the concept of *azimuthal-evanescent modes*. By exploiting the above decomposition, the direct numerical summation of the normal-mode series becomes feasible up to $ka \approx 1000$. For calculations at even higher frequencies ($ka \rightarrow \infty$), asymptotic expressions are derived by using appropriate integral representations of Bessel and Hankel functions, in conjunction with the method of stationary phase. The asymptotic analysis shows that the scattered field is obtained as a superposition of 2-D point sources lying on the boundary of the vertical cylinder, with appropriate amplitudes and phases. Excellent agreement between asymptotic and direct summation numerical results has been demonstrated, at moderate frequencies, where both representations are expected to be valid. © 1997 Acoustical Society of America. [S0001-4966(97)05806-2]

PACS numbers: 43.30.Bp, 43.20.Fn, 43.20.Mv, 43.20.Bi [SAC-B]

INTRODUCTION

Several types of ocean acoustic waveguides, which are encountered in practical situations, especially in environments containing islands and seamounts, are characterized by an abrupt variation of parameters. In these cases it is necessary to resort to the three-dimensional (3-D) Helmholtz equation in order to obtain the field structure around the sharp discontinuity. See, for example, Refs. 1–6.

Exact solutions of 3-D acoustic propagation and scattering problems in the above waveguides, providing a complete description of the pressure field in all spacial coordinates, are of great importance in computational acoustics. Such solutions, except of their own significance in relation with the specific 3-D environment studied, may serve as benchmarks for the assessment of various approximate models developed (or under development) for treating more general environments. In addition, they provide a means for studying the far field, which is important for solving acoustic problems around localized scatterers with more general characteristics, for example, by matching with local elliptic solvers.

In a recent work by Athanassoulis and Prospathopoulos,⁶ a wave theoretic solution based on the Helmholtz equation was developed for the problem of interaction between a source and a vertical cylinder embedded in a plane-horizontal waveguide. The point source is located somewhere outside the cylindrical obstacle, rendering the

problem essentially 3-D. A normal-mode, double-series expansion has been derived for the total acoustic-pressure field outside the cylinder. The solution has been applied to the problem of acoustic scattering from an impenetrable cylindrical island in a range-independent environment, with an arbitrary sound-speed profile.

The series expansion derived in Ref. 6 is an exact analytic solution, which is theoretically valid for all values of the parameters involved in the problem (e.g., for all frequencies, depths, cylinder diameters, etc.). As usually happens to be the case with infinite series representing solutions of wave problems, the numerical efficiency of the direct summation of these series is gradually getting worse as the characteristic wave number increases. The same problem has been clearly demonstrated in Ref. 6 (see, e.g., Fig. 8 of Ref. 6), where it was found, by means of a direct numerical investigation, that the number of azimuthal terms required for the numerical convergence of the total-field series increases rapidly with respect to the frequency when the rest of the parameters are kept constant. This results in prohibitive computational requirements for the calculation of the acoustic field, already from low frequencies and small cylinder diameters.

However, the fact that the double series represents the exact analytical solution of the problem permits, by means of an appropriate analytic and asymptotic treatment, to reformulate the series in forms suitable for calculations in the whole regime of the parameters involved. The purpose of the

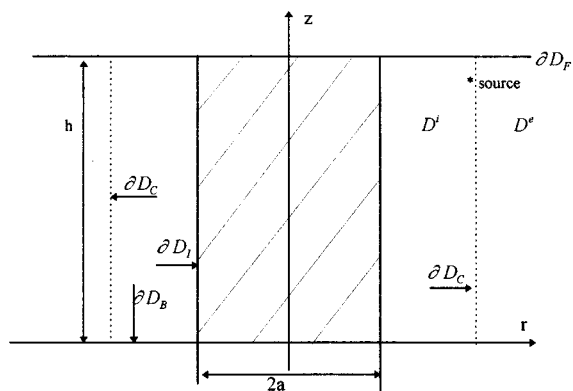


FIG. 1. Geometrical configuration and basic notation.

present work is to carry out such a plan, extending the applicability of the double-series solution, practically to all frequencies. The contents of the present work are organized as follows:

In Sec. I a recapitulation of the main results of Ref. 6, concerning the normal-mode series expansion of the *total acoustic field* is given. In order to improve the rate of convergence of the solution, the total acoustic pressure field is decomposed into the *incident* part, i.e., the axisymmetric unobstructed component of the pressure field, and the *scattered* field. A normal-mode series expansion for the scattered field is derived by applying Graf's addition theorem for Bessel functions of integer order.

In Sec. II the absolute convergence of the scattered-field series is first established. Subsequently, an analytical investigation concerning the rate of convergence of the scattered-field series is presented. It is shown that the essential parameter controlling the number of azimuthal terms required for convergence of the scattered-field series is the nondimensional wave number (ka), based on the radius (a) of the cylinder. The azimuthal terms of the series with index greater than this critical value are shown to decay exponentially. Based on this observation, the concept of *evanescent modes with respect to the azimuthal index* (azimuthal-

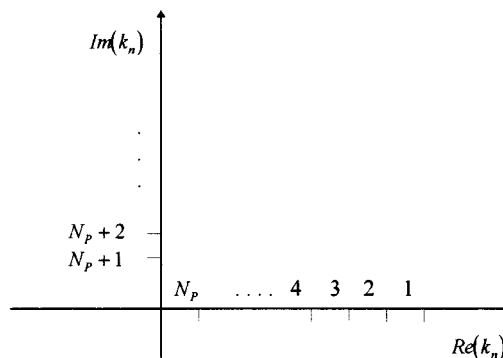


FIG. 2. Ordering of the horizontal wave numbers on the complex k_n plane.

evanescent modes) is introduced. In addition, in the same section, precise information is provided concerning the number of azimuthal-evanescent modes that must be retained in the series in order to obtain numerically convergent results.

The scattered-field series converges considerably faster than the total-field series. Consequently, the decomposition of the field into the incident and scattered components leads to a serious reduction of computational requirements, permitting the straightforward summation of the series and the calculation of the acoustic field for nondimensional wave numbers ka up to 1000. In order to illustrate the usefulness of this approach, numerical results are presented for the problem of underwater acoustic scattering by a cylindrical scatterer ($a = 200$ m) in an isovelocity waveguide, at moderate frequencies ($f = 500$ Hz and $f = 1$ kHz).

For the summation of the scattered-field series at higher frequencies ($ka > 1000$), appropriate asymptotic expressions, obtained by the method of stationary phase, are derived in Sec. III. Both the cases of Dirichlet and Neumann boundary conditions on the surface of the scatterer are considered. The asymptotic analysis shows that the scattered field is obtained as a superposition of 2-D point sources lying on the boundary of the vertical cylinder, with appropriate amplitudes and phases. Numerical results are presented, showing an excellent agreement between the asymptotic results and those obtained by the direct summation of the series, at moderate frequencies, where both representations are expected to be valid.

I. EQUIVALENT REPRESENTATIONS OF THE TOTAL ACOUSTIC FIELD OUTSIDE THE SCATTERER

A. A normal-mode series representation of the total field

Consider the underwater acoustical environment, which is schematically presented in Fig. 1. The environment is determined by a vertical, impenetrable cylinder of radius a , surrounded by a water layer of constant depth h , which is confined between a perfectly reflecting bottom ∂D_B and a pressure-release sea surface ∂D_F . The liquid-cylinder interface is denoted by ∂D_I and the liquid domain is denoted by D . The environment outside the cylinder is assumed to be range independent, with constant density and an arbitrary vertical sound-speed profile. In addition, let $k_0 = \omega/c_0$ denote

the characteristic wave number and c_0 the characteristic value of the sound-speed profile, i.e., $c_0 = \min\{c(z)\}$.

Consider now the boundary value problem for the acoustic field $p(r, \theta, z)$ in D , expressed in cylindrical coordinates; see, e.g., Eqs. (1) and (2) of Ref. 6. The field is generated by a point source, emitting monochromatic sound waves, with angular frequency $\omega = 2\pi f$. Without loss of generality, the point source is assumed to be located at the

point $\mathbf{r}_s = (r_s, 0, z_s) \in D$. The domain D is subdivided by the vertical cylindrical surface ∂D_c , passing through the point source, into the bounded subdomain D^i and the unbounded subdomain D^e . Clearly, $D = D^i \cup D^e \cup \partial D_c$.

In the absence of acoustic energy absorption from the medium or from the boundaries of the waveguide, the *total acoustic pressure field* $p(r, \theta, z)$ admits the following representation:⁶

$$p(r, \theta, z) = \begin{cases} p^i = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} p_{nm}^i = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} (A_{nm} M_{nm}^{(1)} + B_{nm} M_{nm}^{(2)}), & \mathbf{r} = (r, \theta, z) \in D^i, \\ p^e = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} p_{nm}^e = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} C_{nm} M_{nm}^{(1)}, & \mathbf{r} = (r, \theta, z) \in D^e. \end{cases} \quad (1)$$

The functions $M_{nm}^{(q)}(r, \theta, z)$, $q = 1, 2$, denote the normal modes, and are given in terms of Hankel and trigonometric functions as follows:

$$M_{nm}^{(q)}(r, \theta, z) = \tilde{g}_n(z) H_m^{(q)}(k_n r) \cos(m\theta), \quad (2)$$

$$k_n = k_0 \sqrt{1 - \lambda_n^2}, \quad n \in N, \quad (3)$$

where $\{\lambda_n\}_{n=1, \infty}$ is the countably infinite system of eigenvalues of the vertical Sturm–Liouville problem (see, for example, Refs. 6 and 7), and $\{\tilde{g}_n(z) = g_n(z)/\|g_n\|\}_{n=1, \infty}$ is used to symbolize the corresponding set of normalized eigenfunctions (vertical normal modes). In the present case all λ_n^2 are real numbers.^{7,8}

The horizontal wave numbers $\{k_n\}$ are ordered as shown in Fig. 2. The finite subset $\{k_n\}_{n=1, N_p}$ gives rise to the finite number (N_p) of propagating modes, while $\{k_n\}_{n=N_p+1, \infty}$ corresponds to the countably infinite system of evanescent modes. The propagating component of the field is given by truncating the double series, Eq. (1), with respect to the vertical index (n) and retaining the first $n \leq N_p$ terms. For the *isovelocity* case, which shall be used in the present paper for numerical experimentation, N_p is defined as the integer part of the critical value N_{cr} , i.e., $N_p = [N_{cr}]$, given by

$$N_{cr} = \frac{k_0 h}{\pi} + \frac{1}{2} = \frac{1}{2} \left(1 + \frac{f}{f_0} \right), \quad (4)$$

where $f_0 = c_0/4h$ denotes the cutoff frequency of the isovelocity waveguide.

The following three nondimensional wave numbers will enter into play in the course of our analysis: $a_n = k_n a$, $s_n = k_n r_s$, and $r_n = k_n r$. (Only the third wave number, r_n , is dependent on the field point coordinates through the radius r .) The coefficients B_{nm} and C_{nm} of the normal-mode series, Eq. (1), can be expressed in terms of the coefficients A_{nm} by imposing the source conditions on ∂D_c as follows:⁶

$$B_{nm} = -E_n \frac{e_m}{2} H_m^{(1)}(s_n), \quad (5)$$

$$C_{nm} = A_{nm} - E_n \frac{e_m}{2} H_m^{(2)}(s_n), \quad (6)$$

where

$$E_n = -\frac{j}{4} \tilde{g}_n(z_s), \quad j = \sqrt{-1}, \quad (7)$$

and $e_m = 2 - \delta_{m0}$ stands for the Neumann symbol (δ_{nm} is Kronecker's delta).

Finally, the coefficients A_{nm} are determined by imposing the boundary condition on the surface of the scatterer ∂D_I . In the case of a *Dirichlet boundary condition* on ∂D_I ($p = 0$ on ∂D_I , i.e., the case of an acoustically soft island) the analytical result is⁶

$$A_{nm} = E_n \frac{e_m}{2} H_m^{(1)}(s_n) \frac{H_m^{(2)}(a_n)}{H_m^{(1)}(a_n)}, \quad (8)$$

while in the case of *Neumann boundary condition* on ∂D_I ($\partial p / \partial n = 0$ on ∂D_I , i.e., the case of an acoustically hard island) A_{nm} is given by⁶

$$A_{nm} = E_n \frac{e_m}{2} H_m^{(1)}(s_n) \frac{\Phi_m^{(2)}(a_n)}{\Phi_m^{(1)}(a_n)}, \quad (9)$$

where $\Phi_m^{(q)}(x) = [dH_m^{(q)}(x)]/dx$ denotes the derivatives of the corresponding Hankel functions with respect to the argument.

B. Derivation of the normal-mode series representation of the scattered field

The total acoustic field outside the cylinder ∂D_I , Eq. (1), intrinsically contains the unobstructed contribution of the source in the same waveguide, which is otherwise termed the *incident field*. This part is singular at the position of the source ($\mathbf{r} = \mathbf{r}_s$), and, consequently, it is natural to expect that a large number of azimuthal terms is required for the total-field series to converge, especially as frequency increases and as the field point is approaching the source. One way to improve the rate of convergence of the normal-mode series is

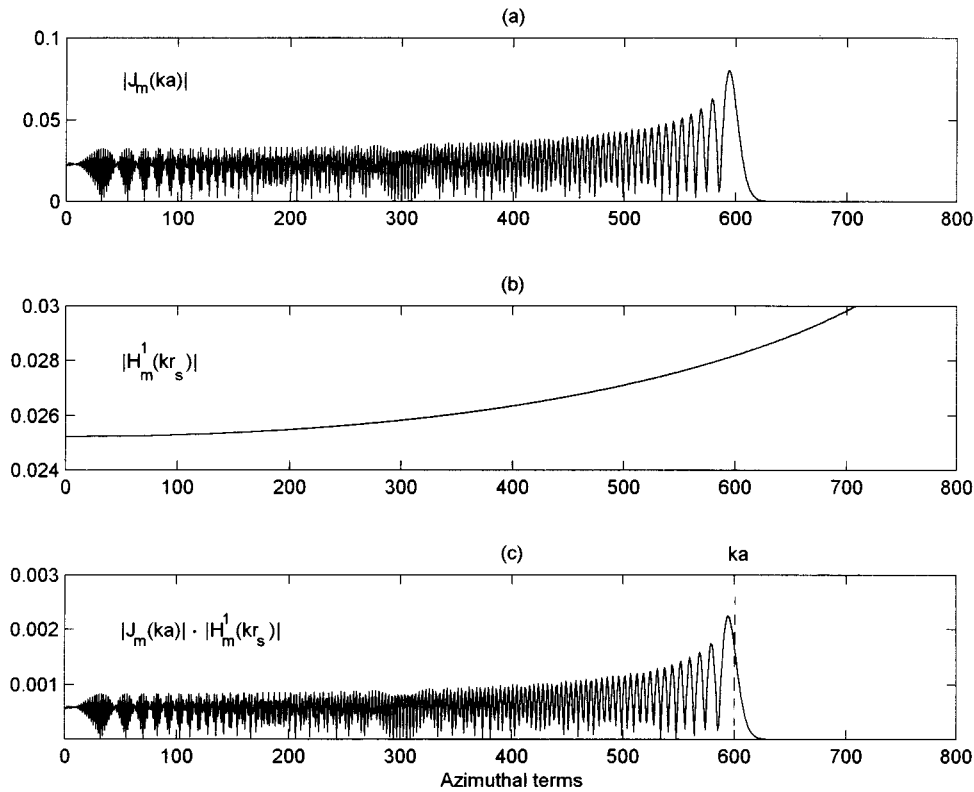


FIG. 3. The concept of the azimuthal-evanescent modes. The values of parameters for this example are: $ka=600$, $kr_s=1000$. (a) Plot of the sequence $\{|J_m(ka)|\}_{m=0}^{800}$; (b) plot of the sequence $\{|H_m^{(1)}(kr_s)|\}_{m=0}^{800}$; (c) plot of the product $\{|J_m(ka) \cdot H_m^{(1)}(kr_s)|\}_{m=0}^{800}$. It can be clearly observed that for $m > M_{cr} = ka = 600$, the product $|J_m(ka) \cdot H_m^{(1)}(kr_s)|$ decays exponentially.

offered by decomposing the total acoustic field into the *incident* and the *scattered* components and to examine further the latter part. The incident field needs not be summed up with respect to the azimuthal index, since it is given by⁸⁻¹⁰

$$p_s(\mathbf{r}) = \frac{j}{4} \sum_{n=1}^{\infty} \tilde{g}_n(z) \tilde{g}_n(z_s) H_0^{(1)}(w_n), \quad (10)$$

where (see also Fig. 1),

$$w_n = k_n w \quad \text{and} \quad w = \sqrt{r_s^2 + r^2 - 2r_s r \cos \theta} \quad (11)$$

is the nondimensional wave number based on the horizontal distance (w) of the field point from the source. To identify the incident field, Eq. (10), as a part of the total field, Eq. (1), use of the following expansion is made:

$$H_0^{(1)}(w_n) = \begin{cases} \sum_{m=0}^{\infty} e_m H_m^{(1)}(s_n) J_m(r_n) \cos m\theta, & \mathbf{r} \in D^i \\ \sum_{m=0}^{\infty} e_m H_m^{(1)}(r_n) J_m(s_n) \cos m\theta, & \mathbf{r} \in D^e \end{cases}, \quad (12)$$

which is referred to in the literature [see, e.g., Ref. 11 (Sec. 11.3) and Ref. 12 (Sec. 9.1.79)] as the Graf's addition theorem of Hankel functions.

The *scattered* field $p_d(\mathbf{r})$ is obtained by subtracting the incident field from the total field,

$$p_d(\mathbf{r}) = p(\mathbf{r}) - p_s(\mathbf{r}). \quad (13)$$

By substituting Eqs. (1), (10), and (12) into Eq. (13), and after some algebra, the following expression, unified for points $\mathbf{r} \in D^i$ and $\mathbf{r} \in D^e$, is derived:

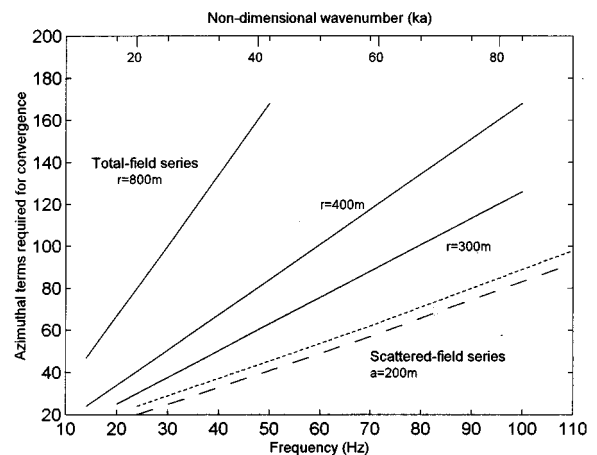


FIG. 4. Number of azimuthal terms versus frequency required to achieve numerical convergence of the initial normal-mode series, Eq. (1), for the case of a soft island (solid lines). The dotted line represents the number of azimuthal terms sufficient for convergence of the scattered-field series, Eq. (14), which is independent from field point coordinates. The dashed line corresponds to estimate (24). Island radius $a=200$ m. Source location $r_s=200$ m. Receiver locations: $\{r_i\} = \{800 \text{ m}, 400 \text{ m}, 300 \text{ m}\}$.

$$p_d(\mathbf{r}) = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} p_{nm}^{(d)} = \sum_{n=1}^{\infty} \tilde{g}_n(z) E_n \sum_{m=0}^{\infty} e_m \Lambda_{nm} H_m^{(1)}(s_n) \times H_m^{(1)}(r_n) \cos m\theta, \quad \mathbf{r} \in D. \quad (14)$$

The coefficients Λ_{nm} of the scattered-field series are given by

$$\Lambda_{nm} = \begin{cases} \Lambda_{nm}^{(s)} = \frac{J_m(a_n)}{H_m^{(1)}(a_n)}, & \text{Dirichlet boundary condition on } \partial D_I, \\ \Lambda_{nm}^{(h)} = \frac{J'_m(a_n)}{\Phi_m^{(1)}(a_n)}, & \text{Neumann boundary condition on } \partial D_I, \end{cases} \quad (15)$$

where $J'_m(x) = dJ_m(x)/dx$.

Accordingly, the total acoustic pressure field in the obstructed waveguide can be calculated by

$$p(r, \theta, z) = \frac{j}{4} \sum_{n=1}^{\infty} \tilde{g}_n(z) \tilde{g}_n(z_s) H_0^{(1)}(w_n) + p_d(r, \theta, z), \quad \mathbf{r} \in D = D^i \cup D^e, \quad (16)$$

where $p_d(r, \theta, z)$ is the scattered field, as provided by Eqs. (14) and (15).

Equations (1) and (16) are mathematically equivalent representations of the total field outside the scatterer. How-

ever, the double series (14) representing $p_d(r, \theta, z)$ exhibits much better convergence properties than the double series, Eq. (1), representing the total field.

II. THE AZIMUTHAL CONVERGENCE OF THE SCATTERED-FIELD SERIES

In this section the improved convergence behavior of the scattered-field series, Eqs. (14) and (15), are established analytically and confirmed numerically. This fact results in significant savings of computational power, thus enabling the direct calculation of the acoustic field well above the limitations encountered when use of the total-field series representation is made.

Since the infinite summation with respect to the vertical index (n) can be truncated to a finite one, retaining only the corresponding propagating modes ($1 \leq n \leq N_p$), the asymptotic behavior of the general term $p_{nm}^{(d)}$ and the convergence properties of the scattered-field series are examined only with respect to the azimuthal index (m).

A. Study of the absolute convergence of the scattered-field series

Observing that the eigenfunctions ($\tilde{g}_n(z)$) of the vertical problem are bounded, and by using the well-known property of Hankel functions, that their modulus, for any fixed order, is a monotonically decreasing function of the argument [see, Ref. 11 (Sec. 13.74)], we derive from Eq. (14) the following bounds for the absolute value of the general term ($p_{nm}^{(d)}$) of the scattered-field series:

$$|p_{nm}^{(d)}| \leq \begin{cases} c_D e_m |J_m(a_n)| |H_m^{(1)}(s_n)|, & \text{Dirichlet b.c. on } \partial D_I \\ c_N e_m \frac{|J'_m(a_n)|}{|\Phi_m^{(1)}(a_n)|} |H_m^{(1)}(s_n)| |H_m^{(1)}(a_n)|, & \text{Neumann b.c. on } \partial D_I \end{cases} \leq \tilde{c} |\tilde{p}_{nm}|, \quad (17)$$

where $|\tilde{p}_{nm}| = |J_m(a_n)| |H_m^{(1)}(s_n)|$, and $m > a_n$.

Quantities c_D , c_N , and \tilde{c} appearing in the above inequalities are positive constants independent from indices n , m , and the field-point coordinates. For the derivation of the second inequality in the case of a Neumann boundary condition on ∂D_I , use is also made of inequality (A5) proved in Appendix A.

In order to establish the convergence properties of the scattered-field series, it suffices to study the *reduced series* with general terms $|\tilde{p}_{nm}|$, defined by Eq. (17). It should be stressed here that, the reduced series are independent from the field-point coordinates (r, θ, z), and thus any results obtained characterize uniformly the absolute convergence of the scattered-field series throughout the whole liquid domain. A first convergence result can be very easily obtained by using the asymptotic forms of Hankel functions for large orders:^{11,12}

$$|\tilde{p}_{nm}| = |J_m(a_n) H_m^{(1)}(s_n)| \leq c_1 |J_m(a_n) Y_m(s_n)|$$

$$\leq c_2 \sqrt{\frac{1}{2\pi m}} \left(\frac{ea_n}{2m}\right)^m \sqrt{\frac{2}{\pi m}} \left(\frac{es_n}{2m}\right)^{-m} \leq c_3 \frac{1}{m} \left(\frac{a}{r_s}\right)^m, \quad m \rightarrow \infty, \quad (18)$$

where c_1 , c_2 , c_3 are positive constants. Since always $a < r_s$ (the point source lies outside the cylinder ∂D_I), estimate (18) shows that the series $\sum_m |\tilde{p}_{nm}|$, and thus the series (14), converge faster than a geometric series with ratio a/r_s . In the next subsection it will be shown that, for each vertical index n , the exponential decay of the general term $|\tilde{p}_{nm}|$ of the reduced series starts to hold immediately after the azimuthal index m has exceeded a critical value M_{cr} .

B. The number of required azimuthal terms for convergence

In order to obtain concrete information concerning the minimum number of azimuthal terms required for the numerical summation of the scattered-field series (14), we now turn to a more detailed investigation of the general term

$|\tilde{p}_{nm}|$ of the reduced series, Eq. (17). A typical plot of $|\tilde{p}_{nm}| = |J_m(a_n)| \cdot |H_m^{(1)}(s_n)|$, considered as a function of the azimuthal index m , is presented in Fig. 3 for $a_n = k_n a = 600$ and $s_n = k_n r_s = 1000$. It can be observed from this figure that $|\tilde{p}_{nm}|$ decays rapidly after the azimuthal index m has exceeded the critical value $M_{cr} = a_n$, which is the argument of the *Bessel-J* function in the product. This suggests that, for each $n = 1, 2, \dots, N_p$, $|\tilde{p}_{nm}|$ is a monotonically decaying sequence of m for $m > M_{cr} = a_n$. This behavior of the general term of the reduced series and, accordingly, of the scattered-field series, is analytically confirmed as follows:

In the interval $a_n < m < s_n$, the functional values of $H_m^{(1)}(s_n)$ remain bounded [$|H_m^{(1)}(s_n)| \leq |H_{s_n}^{(1)}(s_n)|$]; consequently, the behavior of $|\tilde{p}_{nm}|$ is determined by $|J_m(a_n)|$. By considering the following asymptotic expansions of the *Bessel-J* function, in terms of the Airy-Ai function [see, e.g., Ref. 12 (Secs. 9.3.4 and 9.3.23) and Ref. 13 (Sec. 8.455)]

$$J_m(x) \approx \left(\frac{2}{m}\right)^{1/3} \cdot \text{Ai}\left(\frac{2^{1/3}(m-x)}{m^{1/3}}\right), \quad m > x,$$

we conclude

$$|\tilde{p}_{nm}| < C_a \left(\frac{2}{m}\right)^{1/3} \cdot \text{Ai}\left(\frac{2^{1/3}(m-a_n)}{m^{1/3}}\right), \quad \text{for } m > a_n, \quad (19)$$

where C_a is a positive constant. On the other hand, when $m > s_n$, the factor $|H_m^{(1)}(s_n)|$ in the product defining the general term of the reduced series can be approximated by means of the Airy-Bi function as follows:¹²

$$|H_m^{(1)}(y)| \approx \left(\frac{2}{m}\right)^{1/3} \cdot \text{Bi}\left(\frac{2^{1/3}(m-y)}{m^{1/3}}\right), \quad m > y,$$

and thus, the following expression is obtained for the general term of the reduced series for $m > s_n$:

$$|\tilde{p}_{nm}| < C_s \left(\frac{2}{m}\right)^{2/3} \cdot \text{Ai}\left(\frac{2^{1/3}(m-a_n)}{m^{1/3}}\right) \cdot \text{Bi}\left(\frac{2^{1/3}(m-s_n)}{m^{1/3}}\right), \quad \text{for } m > s_n > a_n, \quad (20)$$

where C_s is also a positive constant. To understand better the rate of decay of the right-hand sides of Eqs. (19) and (20) use will be made of the standard asymptotic expansions of Airy functions for large arguments,¹²

$$\text{Ai}(x) \approx \frac{\exp(-2x^{3/2}/3)}{2\pi^{1/2}x^{1/4}} \quad \text{and} \quad (21)$$

$$\text{Bi}(x) \approx \frac{\exp(2x^{3/2}/3)}{\pi^{1/2}x^{1/4}}, \quad \text{for } x \rightarrow \infty,$$

which become very effective even for $x > 2$. Taking Eqs. (21) into account, the following estimates of the general term $|\tilde{p}_{nm}|$ are obtained:

$$|\tilde{p}_{nm}| \leq C \cdot \begin{cases} q_a(m; a_n), & 1 < m/a_n \leq r_s/a, \\ q_s(m; a_n, r_s/a), & 1 < r_s/a < m/a_n, \end{cases} \quad (22)$$

where C is a positive constant and

$$q_a(m; a_n) = \frac{1}{m^{1/2}[(1-a_n)/m]^{1/4}} \cdot \exp\left[-\frac{2^{3/2}}{3} m \left(1 - \frac{a_n}{m}\right)^{3/2}\right], \quad (23a)$$

$$q_s(m; a_n, r_s/a) = \frac{\exp[-(2^{3/2}/3)m(1-a_n/m)^{3/2}\{1 - ([m - (r_s/a)a_n]/(m-a_n))^{3/2}\}]}{m(1-a_n/m)^{1/4}(1-(r_s/a)(s_n/m))^{1/4}}. \quad (23b)$$

The following conclusions can be easily drawn from Eqs. (22) and (23):

(i) As soon as m exceeds a_n the general term $|\tilde{p}_{nm}|$ of the reduced series starts to decay exponentially with respect to m . More precisely, in the interval $a_n < m < s_n = (r_s/a)a_n$, the rate of decay of $|\tilde{p}_{nm}|$ is given by $m^{-1/2} \exp(-D_0 m)$, where D_0 is a constant of order $O(1)$. It is exactly this behavior that is depicted in Fig. 3(c), immediately after m has exceeded the value $a_n = 600$. This behavior continues to hold up to $m = s_n = (r_s/a)a_n$, the interval $[a_n, (r_s/a)a_n]$ being enlarged for higher frequencies (larger a_n), as well as for larger values of the ratio r_s/a . Numerical experience has shown that highly accurate summation of the series is always achieved within this m interval, as far as $r_s/a > 1.1$ and $a_n > 50$. This point will also be discussed below.

(ii) In the interval $m > s_n = (r_s/a)a_n$, the term $|\tilde{p}_{nm}|$ decays also exponentially but, in accordance with Eqs. (22) and (23b), with an ever decreasing rate. Note, however, that

within any fixed m interval of the type; $(r_s/a)a_n < m < la_n$, where l is a fixed number greater than r_s/a , the rate of decay of $|\tilde{p}_{nm}|$ is again of the form $m^{-1} \exp(-D_l m)$, where D_l is a constant of order $O(1)$ dependent on l , but smaller than D_0 . From the numerical point of view this interval is not important, since the numerical convergence, if not already accomplished in the region $a_n < m < (r_s/a)a_n$, necessitates only a very few additional terms in the region $m > s_n = (r_s/a)a_n$.

Based on the above considerations, we arrive at the following conclusions:

(i) For each $n = 1, 2, \dots, N_p$, the general terms $|\tilde{p}_{nm}|$ of the reduced series behaves asymptotically ($m \rightarrow \infty$) as $m^{-1}(a/r_s)^m$. What is more important is, that the exponential decay of $|\tilde{p}_{nm}|$ starts to hold immediately after the azimuthal index m has exceeded the critical value $M_{cr} = a_n = k_n a$. This fact ensures an efficient numerical summation of the series $\sum_m |\tilde{p}_{nm}|$ by con-

sidering only a few additional terms above $m = a_n$.

(ii) The set of normal modes appearing in the representation of the scattered field, Eqs. (14) and (15), is subdivided into the classes of propagating and evanescent modes, with respect to both the vertical index n and the azimuthal index m . The transition from vertical-propagating to vertical-evanescent modes is determined by $N_p = [N_{cr}]$, given by Eq. (4). For each horizontal wave number k_n , the transition from azimuthal-propagating to azimuthal-evanescent modes is determined by $M_{cr} = a_n = k_n a$.

Since $k_1 > k_2 > \dots > k_{N_p}$, the number of required azimuthal terms in order to achieve numerical convergence of the scattered-field series must be greater than $a_1 = k_1 \cdot a$. For the *isovelocity waveguide* we have

$$k_1 a = \sqrt{1 - \left(\frac{f_0}{f}\right)^2} \frac{2\pi f a}{c_0}, \quad (24)$$

where f_0 is the cutoff frequency.

In Fig. 4 the number of azimuthal terms required for the efficient numerical summation of series (1) is compared to estimate (24). In this figure, which is partially reproduced

from Ref. 6, the number of necessary azimuthal terms for series (1) to converge is presented vs frequency (solid lines) and for various field-point radii, as estimated by direct numerical investigation (i.e., by systematically varying the number of azimuthal terms retained in the series until numerical convergence is achieved). The studied environment, denoted by Π , is characterized by the following parameters:

$$\Pi = \{h = 250 \text{ m}, a = 200 \text{ m}, z_s = 2h/3, r_s = 2000 \text{ m}, c(z) = c_0 = 1500 \text{ m/s}\}. \quad (25)$$

Predictions based on Eq. (24) are presented in the same figure by using a dashed line. Since, however, Eq. (24) provides just a lower bound for the number of azimuthal terms required for the convergence of the scattered-field series, it is expedient to continue our investigation in order to determine the number of azimuthal-evanescent modes that are sufficient for a highly accurate numerical summation. Recalling that the reduced series, Eq. (17), depends on two parameters, specifically the characteristic nondimensional wave number $a_1 = k_1 a$ and the ratio r_s/a , and denoting by $S(M) = S(M; k_1 a, r_s/a)$ the following partial sums:

$$S(M; k_1 a, r_s/a) = \begin{cases} \sum_{m=0}^M e_m |J_m(k_1 a)| |H_m^{(1)}(k_1 r_s)|, & \text{Dirichlet b.c. on } \partial D_I, \\ \sum_{m=0}^M e_m \frac{|J'_m(k_1 a)|}{|\Phi_m^{(1)}(k_1 a)|} |H_m^{(1)}(k_1 r_s)| |H_m^{(1)}(k_1 a)|, & \text{Neumann b.c. on } \partial D_I, \end{cases} \quad (26)$$

the number of the *sufficient* terms for the azimuthal convergence of the scattered field can be determined by comparing $S(M; k_1 a, r_s/a)$ with $S_{\text{inf}} = S(M \rightarrow \infty; k_1 a, r_s/a)$, for various values of the parameters. In Fig. 5 the quantity $S(M; k_1 a, r_s/a)/S_{\text{inf}}$, providing the relative error due to the truncation of the series (26), is presented versus the number of the retained azimuthal-evanescent terms $(M - k_1 a)$. The results shown in this figure correspond to the following values of parameters $k_1 a = 100$ and 1000 , and $r_s/a = 1.1, 2$, and 10 . The line corresponding to an error of 1% is also plotted. It can be drawn from this figure that, in all cases of interest, a small number (5–10) of azimuthal-evanescent terms is sufficient for obtaining accurate results. The dotted line in Fig. 4, appearing just above the dashed line corresponding to Eq. (24), represents the exact number of terms that are sufficient for the highly accurate calculation of the acoustic field (error < 1%) by means of Eq. (16). It is clear that the computational requirements for the calculation of the acoustic field by using Eq. (16), as compared to the ones of Eq. (1), are drastically reduced. It should be stressed once again here, that the number of required azimuthal terms for the scattered-field series to converge is independent from the field-point coordinates.

C. Numerical results

As previously established, the alternative representation, Eq. (16), of the acoustic pressure field is much better suited for numerical calculations than the initial representation, Eq. (1). As a result, calculations are enabled well above the limitations encountered when using the expansion (1).

An example of application of the scattered-field series to the calculation of the acoustic pressure field around a vertical cylinder is presented in Figs. 6 and 7. In these figures, 2-D patterns of the transmission loss (coherent TL in dB), at the depth of the source and for two frequencies, $f = 500$ Hz and $f = 1000$ Hz, are shown. The parameters defining the acoustic environment are kept the same as defined by Eq. (25). The case of a Dirichlet boundary condition on the surface of the scatterer (soft cylindrical island) is presented in Fig. 6, and the case of a Neumann boundary condition (hard cylindrical island) is presented in Fig. 7. As frequency increases, the edge of the shadow becomes more distinct at the rear of the island, and a trend can be observed for this boundary to follow a smooth curve. The abrupt variation of TL toward increased losses at the rear of the cylindrical island is much steeper in the case of a soft island. Outside the angular section defined by the tangents drawn from the source to the island, and away from the liquid–solid interface, the incident component is dominant, as expected.

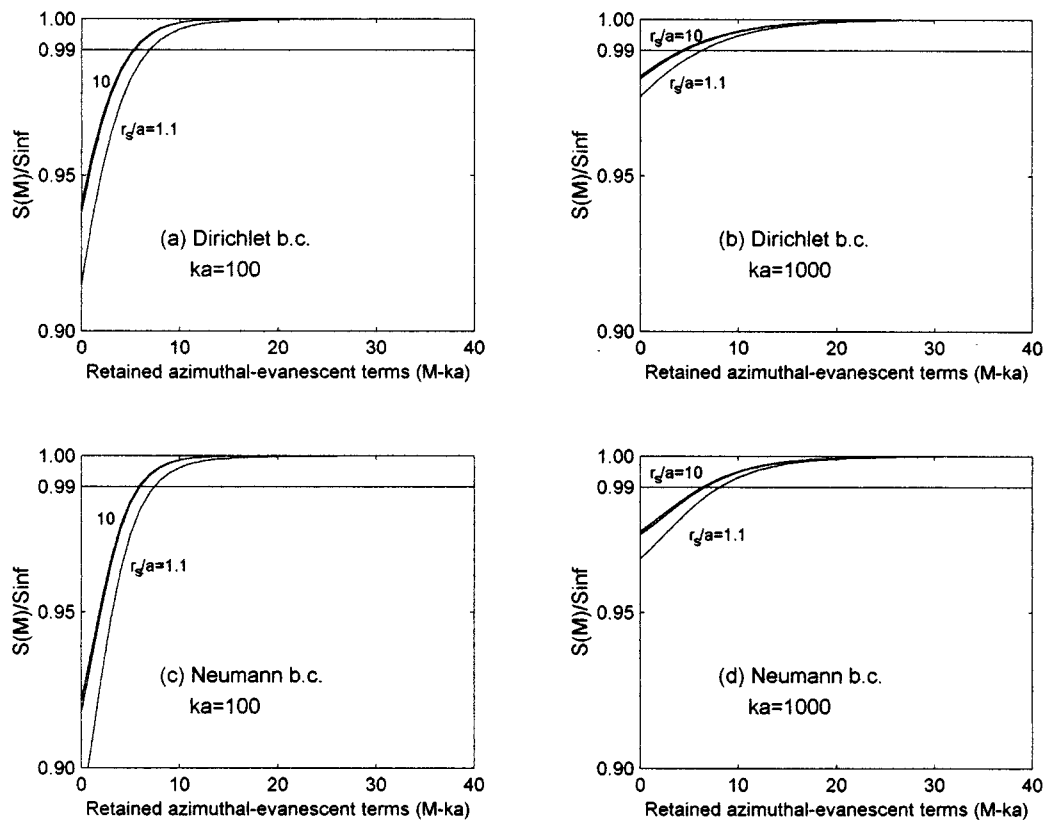


FIG. 5. Study of the number of azimuthal-evanescent modes which must be retained in the scattered-field series in order to obtain numerical convergence. Plot of the functions $S(M;ka,r_s/a)/S_{\text{inf}}$, Eq. (26), vs the number of the azimuthal-evanescent terms ($M-ka$) retained in the series. (a) Dirichlet boundary condition, $r_s/a = 1.2, 2, 10$, and $ka = 100$; (b) The same as (a) but for $ka = 1000$; (c) Neumann boundary condition, $r_s/a = 1.1, 2, 10$ and $ka = 100$; (d) the same as (c) but for $ka = 1000$. (Note that the curves corresponding to $r_s/a = 2$ and $r_s/a = 10$ are almost coincident.)

III. HIGH-FREQUENCY ASYMPTOTIC EXPANSIONS OF THE SCATTERED FIELD

For the summation of the scattered-field series at even higher values of the characteristic nondimensional wave number $k_0a = a\omega/c_0$, another, more appropriate asymptotic procedure should be used. In this connection, in the present section the azimuthal series is first represented by a suitable integral form and, subsequently, the principle of stationary

phase is applied. To start with, consider the part of the scattered field, Eq. (14), corresponding to the propagating horizontal wave numbers $\{k_n\}_{n=1,N_p}$. This part can be equivalently written in the form

$$p_d(\mathbf{r}) = \sum_{n=1}^{N_p} E_n \tilde{g}_n(z) f_n(r, \theta), \quad (27)$$

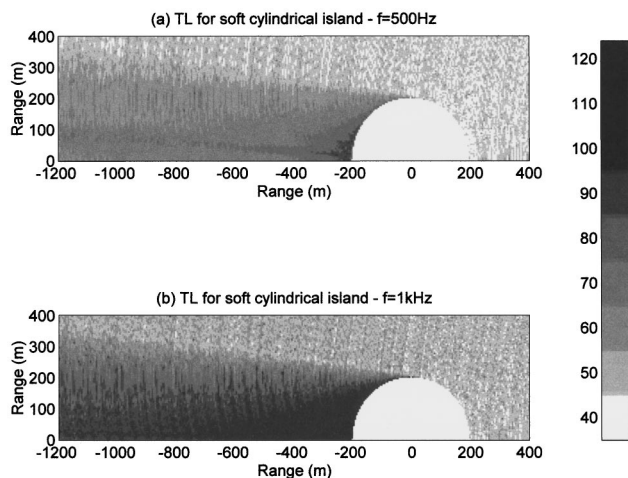


FIG. 6. Transmission loss (TL) calculated by means of the scattered-field series, Eqs. (14) and (16), in the case of a Dirichlet boundary condition on the surface of the cylinder (soft scatterer). (a) $f = 500$ Hz. (b) $f = 1$ kHz.

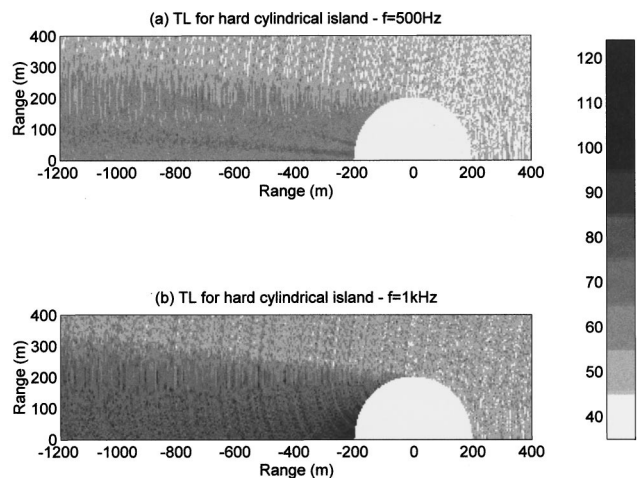


FIG. 7. Transmission loss (TL) calculated by means of the scattered-field series, Eqs. (14) and (16), in the case of a Neumann boundary condition on the surface of the cylinder (hard scatterer). (a) $f = 500$ Hz. (b) $f = 1$ kHz.

where E_n is a constant given by Eq. (7), and the functions $f_n(r, \theta)$ are defined as follows:

$$f_n(r, \theta) = \sum_{m=0}^{\infty} e_m \cos m\theta \Lambda_{nm} H_m^{(1)}(s_n) H_m^{(1)}(r_n). \quad (28)$$

In accordance with Eq. (27), the scattered field at every

depth, $z = \text{const.}$, can be considered as a weighted linear superposition of the 2-D fields $f_n(r, \theta)$, each one corresponding to the horizontal wave number k_n . Using the integral representation of products of Bessel and Hankel functions of integer order, detailed in Appendix B, the following alternative expression for the functions $f_n(r, \theta)$ is obtained:

$$2\pi f_n(r, \theta) = \begin{cases} \int_{\varphi=0}^{\varphi=\pi} H_0^{(1)}[k_n Q(\varphi)] \{F_n(\theta + \varphi) + F_n(\theta - \varphi)\} d\varphi, & \text{Dirichlet b.c.,} \\ \int_{\varphi=0}^{\varphi=\pi} H_1^{(1)}[k_n Q(\varphi)] \frac{r \cos \varphi - a}{Q(\varphi)} \{F_n(\theta + \varphi) + F_n(\theta - \varphi)\} d\varphi, & \text{Neumann b.c.,} \end{cases} \quad (29)$$

where $Q(\varphi) = \sqrt{r^2 + a^2 - 2ra \cos \varphi}$ and,

$$F_n(\varphi) = M_n(\varphi) \exp(jA_n(\varphi)) \\ \equiv \begin{cases} \sum_{m=0}^{\infty} e_m \frac{H_m^{(1)}(s_n)}{H_m^{(1)}(a_n)} \cos(m\varphi), & \text{Dirichlet b.c.,} \\ \sum_{m=0}^{\infty} e_m \frac{H_m^{(1)}(s_n)}{\Phi_m^{(1)}(a_n)} \cos(m\varphi), & \text{Neumann b.c.} \end{cases} \quad (30)$$

Let it be noted here that the functions $F_n(\varphi)$ of the azimuthal argument φ are parametrically dependent on the corresponding horizontal (nondimensional) wave number $a_n = k_n a$, and the relative position of the source $s_n/a_n = r_s/a$, i.e.,

$$F_n\left(\varphi; a_n, \frac{r_s}{a}\right) = M_n\left(\varphi; a_n, \frac{r_s}{a}\right) \exp\left[jA_n\left(\varphi; a_n, \frac{r_s}{a}\right)\right]. \quad (31)$$

On the other hand, in contradistinction to the functions $f_n(r, \theta)$ defined by Eq. (28), $F_n(\varphi)$ are entirely independent from the field-point coordinates (r, θ) .

To obtain a better idea of what $F_n(\varphi)$ actually represents, let us consider the horizontal distance of the source from the origin (r_s), that appears as a parameter in Eqs. (30) and (31), to be variable. Then, each $F_n(\varphi)$ is, actually, the total 2-D acoustic field, induced at the field point (r_s, φ) , from a 2-D source located on the (soft/hard) boundary at the

point $(a, 0)$, which radiates energy with a frequency corresponding to the horizontal wave numbers k_n ; cf. the analysis of Ursell concerning creeping modes in a shadow, in Ref. 14. Accordingly, by Eq. (29), $f_n(r, \theta)$ is the 2-D field which is produced by a continuous distribution of source singularities on the circle $r = a$, radiating energy at a nondimensional wave number k_n , with intensity (amplitude and phase) defined by $F_n(\varphi)$. This statement can be considered as the appropriate form of *Huygens principle* for the problem under study.

The expansion of $F_n(\varphi)$ into the exponential form $M_n(\varphi) \exp[jA_n(\varphi)]$ is introduced here because, in this way, the rapidly oscillating function $F_n(\varphi)$ is separated into a nonoscillating modulus and a nonoscillating (rapidly increasing) phase (cf. WKB-approximation and ray theory¹⁵). Representative plots of the modulus $M_n(\varphi)$ and the derivative of the phase function $A_n'(\varphi)$ are shown in Figs. 8 and 9, in the case of a Dirichlet and a Neumann boundary condition on ∂D_I , respectively. In these figures the parameter $a_n = k_n a$ is varied from 100 to 1000 and the ratio r_s/a from 2 to 10. It can be observed from these figures that, for large values of $k_n a$ and a fixed ratio r_s/a , the functions $M_n(\varphi)$ and $A_n'(\varphi)$ obey, approximately, the following “scaling” laws:

$$M_n(\varphi)/M_l(\varphi) \approx \sqrt{k_n/k_l}, \quad A_n'(\varphi)/A_l'(\varphi) \approx k_n/k_l. \quad (32)$$

For $k_n(r-a) \gg 1$, the Hankel functions appearing in the integrands at the right-hand side of Eq. (29) can be approximated by their asymptotic forms. Consequently,

$$f_n(r, \theta) \approx \begin{cases} \frac{1-j}{2\sqrt{\pi^3}} \int_{\varphi=0}^{\varphi=\pi} \frac{\exp[jk_n Q(\varphi)]}{\sqrt{k_n Q(\varphi)}} \{F_n(\theta + \varphi) + F_n(\theta - \varphi)\} d\varphi, & \text{Dirichlet b.c.,} \\ \frac{1+j}{2\sqrt{\pi^3}} \int_{\varphi=0}^{\varphi=\pi} \frac{\exp[jk_n Q(\varphi)]}{\sqrt{k_n Q(\varphi)}} \frac{r \cos \varphi - a}{Q(\varphi)} \{F_n(\theta + \varphi) + F_n(\theta - \varphi)\} d\varphi, & \text{Neumann b.c.} \end{cases} \quad (33)$$

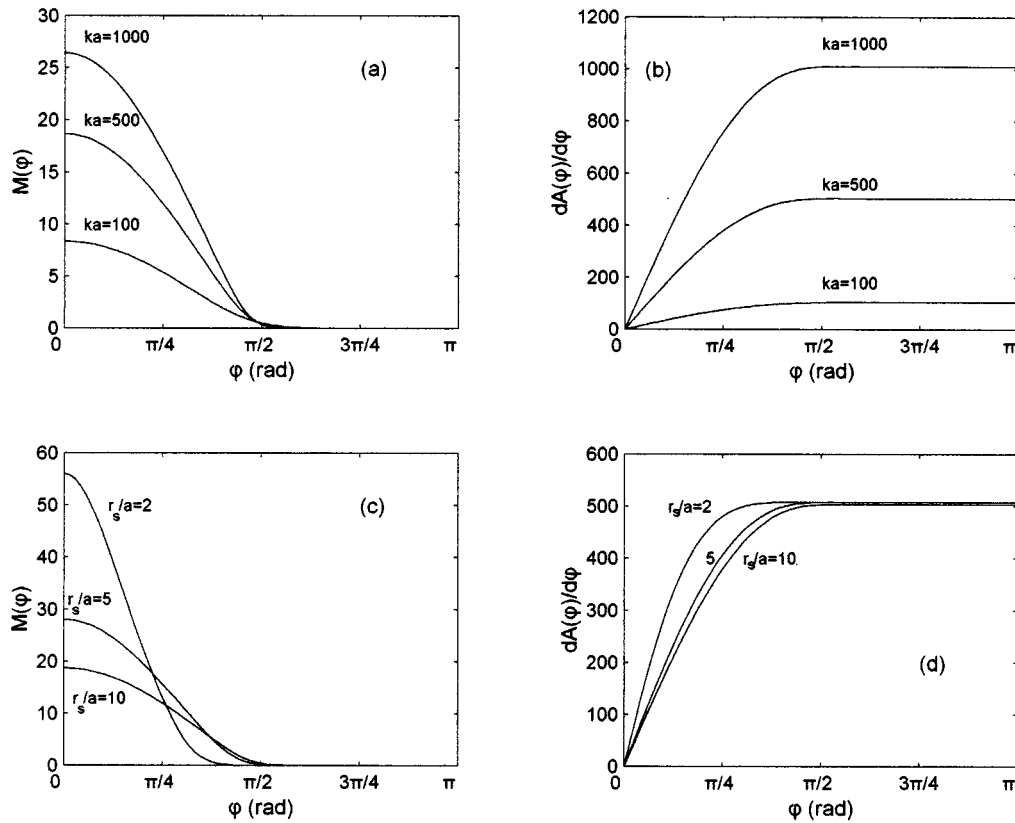


FIG. 8. Modulus $M_n(\varphi)$ and phase derivative $A'_n(\varphi)$ of the function $F_n(\varphi; a_n, r_s/a)$, in the case of a Dirichlet boundary condition on the surface of the cylinder. (a) Plot of $M_n(\varphi; a_n, r_s/a)$ vs the azimuthal angle φ , for $r_s/a=10$ ($a_n=ka=100, 500, 1000$); (b) The same as (a) for $A'_n(\varphi; a_n, r_s/a)$; (c) Plot of $M_n(\varphi; a_n, r_s/a)$ vs φ , for $a_n=ka=500$ ($r_s/a=2, 5, 10$); (d) The same as (c) for $A'_n(\varphi; a_n, r_s/a)$.

Since in the above expressions $k_n Q(\varphi) \gg k_n(r-a) \gg 1$, the involved integrals can be asymptotically evaluated by the method of stationary phase (see, e.g., Refs. 16 and 17). In the case of a *Dirichlet boundary condition* on ∂D_I , we obtain

$$f_n(r, \theta) = \frac{1}{\pi} \frac{M_n(\theta - \varphi_n)}{\sqrt{k_n Q''(\varphi_n) + A''(\theta - \varphi_n)}} \times \frac{\exp[j(k_n Q(\varphi_n) + A_n(\theta - \varphi_n))]}{\sqrt{k_n Q(\varphi_n)}}, \quad (34)$$

where φ_n is the unique root of the algebraic equation,

$$k_n Q'(\varphi) = A'_n(\theta - \varphi), \quad n = 1, 2, \dots \quad (35)$$

In the high-frequency case $k_0 a \gg 1$, the condition $k_n a \gg 1$ is also satisfied by the majority of the propagating modes. Then, from Eqs. (32), (34), and (35), it is deduced that

$$\varphi_n \approx \varphi_1 \quad \text{and} \quad |f_n(r, \theta)| \propto \frac{1}{\sqrt{k_n Q(\varphi_n)}}. \quad (36)$$

Accordingly, the contribution, $f_n(r, \theta)$, from each mode to the scattered field $p_d(r, \theta, z)$ is given as the induced pressure

from a source located at the point (a, φ_n) on the boundary ∂D_I , radiating energy with an amplitude and phase given by Eq. (34). See also Fig. 10.

In the case of a *Neumann boundary condition* on ∂D_I , we obtain from Eq. (33), by expanding the cosine term appearing in the integrand into its equivalent exponential form,

$$\begin{aligned} f_n(r, \theta) = & -\frac{1+j}{2\sqrt{\pi^3}} \int_{\varphi=0}^{\varphi=\pi} \frac{\exp[jk_n Q(\varphi)]}{\sqrt{k_n Q(\varphi)}} \frac{a}{Q(\varphi)} \\ & \times \{F_n(\theta + \varphi) + F_n(\theta - \varphi)\} d\varphi \\ & + \frac{1+j}{4\sqrt{\pi^3}} \int_{\varphi=0}^{\varphi=\pi} \frac{\exp[jk_n Q(\varphi) + j\varphi]}{\sqrt{k_n Q(\varphi)}} \frac{r}{Q(\varphi)} \\ & \times \{F_n(\theta + \varphi) + F_n(\theta - \varphi)\} d\varphi \\ & + \frac{1+j}{4\sqrt{\pi^3}} \int_{\varphi=0}^{\varphi=\pi} \frac{\exp[jk_n Q(\varphi) - j\varphi]}{\sqrt{k_n Q(\varphi)}} \frac{r}{Q(\varphi)} \\ & \times \{F_n(\theta + \varphi) + F_n(\theta - \varphi)\} d\varphi. \end{aligned} \quad (37)$$

All three integrals in the above formula have points of stationarity; therefore, in the case of Neumann boundary condition on the surface of the scatterer, we obtain the following stationary phase approximation for $f_n(r, \theta)$:

$$f_n(r, \theta) = \frac{jk_n a}{\pi} \frac{M_n(\theta - \varphi_{1n})}{\sqrt{k_n Q''(\varphi_{1n}) + A''(\theta - \varphi_{1n})}} \frac{\exp[j(k_n Q(\varphi_{1n}) + A_n(\theta - \varphi_{1n}))]}{[k_n Q(\varphi_{1n})]^{3/2}} \\ - \frac{jk_n r}{2\pi} \frac{M_n(\theta - \varphi_{2n})}{\sqrt{k_n Q''(\varphi_{2n}) + A''(\theta - \varphi_{2n})}} \frac{\exp[j(k_n Q(\varphi_{2n}) + \varphi_{2n} + A_n(\theta - \varphi_{2n}))]}{[k_n Q(\varphi_{2n})]^{3/2}} \\ - \frac{jk_n r}{2\pi} \frac{M_n(\theta - \varphi_{3n})}{\sqrt{k_n Q''(\varphi_{3n}) + A''(\theta - \varphi_{3n})}} \frac{\exp[j(k_n Q(\varphi_{3n}) - \varphi_{3n} + A_n(\theta - \varphi_{3n}))]}{[k_n Q(\varphi_{3n})]^{3/2}}, \quad (38)$$

where $\{\varphi_{1n}, \varphi_{2n}, \varphi_{3n}\}$ are the roots of the algebraic equations,

$$\left\{ \begin{array}{l} \varphi_1: k_n Q'(\varphi_1) = A'_n(\theta - \varphi_1) \\ \varphi_2: k_n Q'(\varphi_2) + 1 = A'_n(\theta - \varphi_2) \\ \varphi_3: k_n Q'(\varphi_3) - 1 = A'_n(\theta - \varphi_3) \end{array} \right\}, \quad n = 1, 2, \dots \quad (39)$$

Therefore, in the high-frequency case, $k_n a \gg 1$, the contribution $f_n(r, \theta)$, of each mode to the scattered-field $p_d(r, \theta, z)$, is given as the induced pressure from the combination of three sources located at the nearby points (a, φ_{1n}) , (a, φ_{2n}) , (a, φ_{3n}) on the boundary ∂D_I , radiating energy with an amplitude and phase given by Eq. (38). See also Fig. 11.

Calculations of the scattered field obtained by using the

asymptotic formulas, Eqs. (34) and (38), and by the direct summation of the scattered-field series, Eqs. (14) and (15), are comparatively presented in Fig. 12. Both the cases of Dirichlet and Neumann boundary conditions on the surface of the cylinder are considered. The parameters defining the acoustical environment are kept the same as given by Eq. (25). The frequency of the source has been selected to be 500 Hz ($ka \approx 420$) in this example, in order that both representations to be valid. Numerical results obtained by the direct and the asymptotic summation of the scattered-field series, presented on the cylindrical surface $r = 1.5a$ and at the level of the source $z = z_s$, are in very good agreement. This result is to be considered as a strong indication of the usefulness and the applicability of the present asymptotic technique to high frequencies.

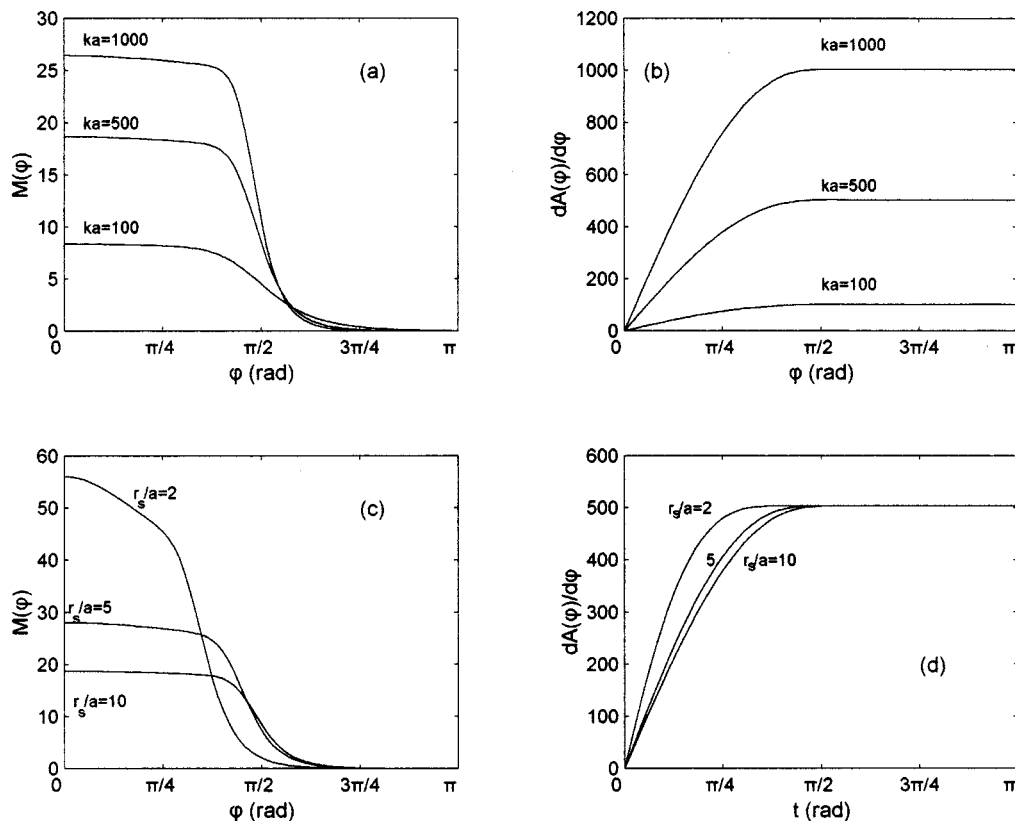


FIG. 9. Modulus $M_n(\varphi)$ and phase derivative $A'_n(\varphi)$ of the function $F_n(\varphi; a_n, r_s/a)$, in the case of a Neumann boundary condition on the surface of the cylinder (a) Plot of $M_n(\varphi; a_n, r_s/a)$ vs the azimuthal angle φ , for $r_s/a = 10$ ($a_n = k_n a = 100, 500, 1000$); (b) The same as (a) for $A'_n(\varphi; a_n, r_s/a)$; (c) Plot of $M_n(\varphi; a_n, r_s/a)$ vs φ , for $a_n = k_n a = 500$ ($r_s/a = 2, 5, 10$); (d) The same as (c) for $A'_n(\varphi; a_n, r_s/a)$.

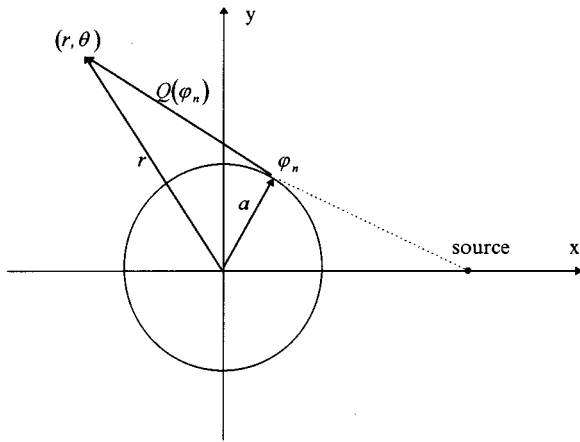


FIG. 10. Geometrical interpretation of the contribution of a normal mode to the scattered field at high frequencies, in the case of a Dirichlet boundary condition on the surface of the cylinder.

IV. CONCLUDING REMARKS

In the present work, the 3-D acoustic scattering from a vertical impenetrable cylinder in a plane-horizontal waveguide has been studied, in the context of normal-mode theory. The analytical solution of the problem, for a Dirichlet or a Neumann boundary condition on the surface of the scatterer, has been derived recently in Ref. 6, in the form of a double-infinite series of normal modes representing the total

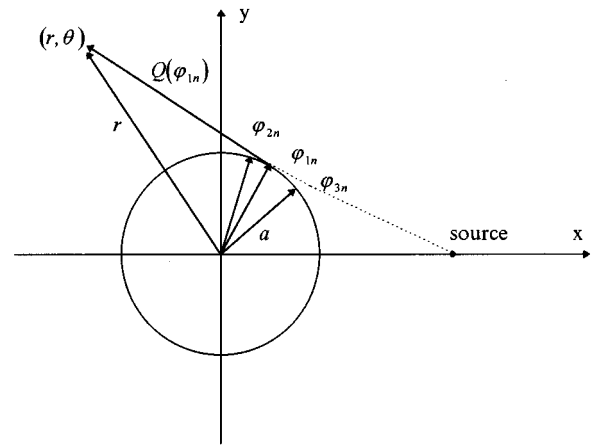


FIG. 11. Geometrical interpretation of the contribution of a normal mode to the scattered field at high frequencies, in the case of a Neumann boundary condition on the surface of the cylinder.

acoustic field. In order to extend the applicability of this solution to higher frequencies, the total field is decomposed into the incident and the scattered parts, and a series expansion of the scattered field is derived. The latter converges considerably faster than the corresponding total-field series, permitting the efficient numerical summation of the normal-mode series up to moderate frequencies.

The main parameter controlling the azimuthal convergence of the scattered-field series is the nondimensional

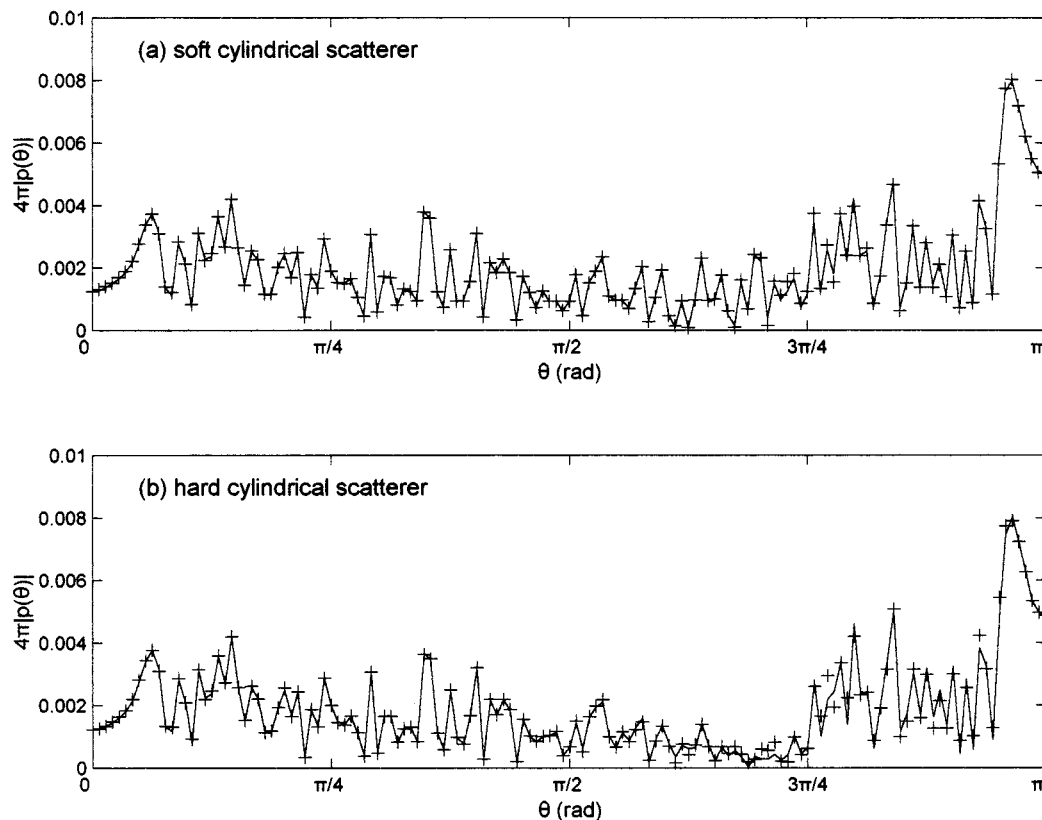


FIG. 12. Calculated scattered pressure modulus by means of direct summation of the scattered-field series, Eq. (14), (solid lines), and by means of the asymptotic formulae Eqs. (27), (34), and (38), (crosses). Source frequency $f=500$ Hz, ($ka \approx 420$). The receiver lies on the circular arc $\{r=1.5a, 0 \leq \theta \leq \pi, z=z_s\}$. (a) Dirichlet boundary condition on the surface of the cylinder; (b) Neumann boundary condition on the surface of the cylinder.

wave number $k_1 a$, based on the radius a of the vertical cylinder. Furthermore, it has been found that the general term of the scattered-field series exhibits an exponential decay, as soon as m has exceeded the value $k_1 a$. This fact justifies the introduction of the concept of *azimuthal-evanescent modes*, and, from the numerical point of view, enables an efficient summation up to $k_1 a \approx 1000$. It is also demonstrated that, for all cases of interest, a small number (5–10) of azimuthal-evanescent terms has to be retained in order to calculate the scattered field with high accuracy.

For acoustic-field calculations at even higher frequencies, asymptotic expressions are derived by using appropriate integral representations of products of Bessel and Hankel functions in conjunction with the method of stationary phase. Numerical results are presented, showing an excellent agreement between the asymptotic results and the direct summation of the series, at moderate frequencies, where both representations are expected to be valid. Thus by combining the two approaches, we have the possibility to obtain an efficient calculation of the acoustic field, using the normal-mode series, for the whole frequency range. This fact is reflected in the somewhat unusual term ‘‘all-frequency normal-mode solution’’ appearing in the title of the present work.

The present solution and especially its high-frequency asymptotics is expected to be useful, in conjunction with boundary integral equation techniques, cf. Ref. 4, in treating acoustic boundary value problems of the same kind around localized scatterers with more complex geometry.

ACKNOWLEDGMENT

The present work was partially supported by the Hellenic Navy.

APPENDIX A: A USEFUL INEQUALITY CONCERNING BESSEL AND HANKEL FUNCTIONS

By applying triangle’s inequality to the recursive formulas defining the derivative of *Bessel-J* functions with respect to the argument [see Ref. 12 (§9.1.27)], we obtain

$$|J'_m(x)| \leq |J_{m+1}(x)| + \frac{m}{x} |J_m(x)|. \quad (\text{A1})$$

By using the fact that, the absolute value of the *Bessel-J* function ($|J_m(x)|$) is a decreasing function of the order (m), for any fixed value of the argument ($x \in R^+$) and for $m > x$ [see also Eq. (19)], we obtain from Eq. (A1),

$$|J'_m(x)| < \left(1 + \frac{m}{x}\right) |J_m(x)|, \text{ for } m > x. \quad (\text{A2})$$

By applying triangle’s inequality to the recursive formulae defining the derivative of *Hankel* functions with respect to the argument,¹² we obtain,

$$\frac{m}{x} |H_m^{(1)}(x)| \leq |\Phi_m^{(1)}(x)| + |H_{m-1}^{(1)}(x)|. \quad (\text{A3})$$

The modulus of the Hankel function $|H_m^{(1)}(x)|$ is a monotonically increasing function of the order (m), for any fixed value of the argument ($x \in R^+$), as it can be easily deduced

by Nicholson’s integral formula; see Ref. 11 (§13.73). Consequently, we obtain from Eq. (A3)

$$|\Phi_m^{(1)}(x)| \geq \left(\frac{m}{x} - 1\right) |H_m^{(1)}(x)|, \text{ for } m > x. \quad (\text{A4})$$

By combining Eqs. (A2) and (A4), we can easily derive the following inequality,

$$\frac{|J'_m(x)|}{|\Phi_m^{(1)}(x)|} < \frac{(m+x)}{(m-x)} \frac{|J_m(x)|}{|H_m^{(1)}(x)|} \leq c \frac{|J_m(x)|}{|H_m^{(1)}(x)|},$$

for $m > x$, (A5)

where c is a positive constant. This inequality is used in Sec. II for the establishment of the absolute convergence of the scattered-field series, in the case of Neumann boundary condition on the surface of the cylinder.

APPENDIX B: INTEGRAL REPRESENTATIONS OF PRODUCTS OF BESSEL AND HANKEL FUNCTIONS OF INTEGER ORDER

From the representation of Hankel functions by means of Graf–Fourier series,^{11,12}

$$\begin{aligned} H_0^{(1)}(kQ(\varphi)) &= \sum_{m=0}^{\infty} e_m J_m(ka) H_m^{(1)}(kr) \cos(m\varphi) \\ &= \sum_{m=0}^{\infty} e_m b_m \cos(m\varphi), \end{aligned} \quad (\text{B1})$$

where $Q(\varphi) = \sqrt{r^2 + a^2 - 2ra \cdot \cos \varphi}$, it follows directly (see also Refs. 1 and 18)

$$\begin{aligned} b_m &= J_m(ka) H_m^{(1)}(kr) \\ &= \frac{1}{\pi} \int_{\varphi=0}^{\pi} H_0^{(1)}(kQ(\varphi)) \cos(m\varphi) d\varphi. \end{aligned} \quad (\text{B2})$$

Differentiating the above series with respect to ka , we get immediately:

$$\begin{aligned} J'_m(ka) H_m^{(1)}(kr) &= \frac{1}{\pi} \int_{\varphi=0}^{\pi} \Phi_1^{(1)}(kQ(\varphi)) \\ &\quad \times \frac{a-r \cos(\varphi)}{Q(\varphi)} \cos(m\varphi) d\varphi, \\ &= \frac{1}{\pi} \int_{\varphi=0}^{\pi} H_1^{(1)}(kQ(\varphi)) \\ &\quad \times \frac{r-a \cos(\varphi)}{Q(\varphi)} \cos(m\varphi) d\varphi. \end{aligned} \quad (\text{B3})$$

Equations (B2) and (B3) define the integral representations which are used in Sec. III for the derivation of high-frequency asymptotic expressions concerning the scattered field.

¹M. J. Buckingham, ‘‘Theory of acoustic propagation around a conical seamount,’’ J. Acoust. Soc. Am. **80**, 256–277 (1986).

²M. J. Taroudakis, G. A. Athanassoulis, and J. P. Ioannidis, ‘‘A variational principle for underwater acoustic propagation in a three-dimensional ocean environment,’’ J. Acoust. Soc. Am. **88**, 1515–1522 (1990).

- ³R. B. Evans, "Three dimensional acoustic scattering from a cylindrical inclusion in a waveguide," *Comput. Acoust.* **2** (1990).
- ⁴J. A. Fawcett, "An efficient three-dimensional boundary integral equation for solving azimuthally symmetric scattering problems in the oceanic waveguide," *J. Acoust. Soc. Am.* **94**, 2307–2314 (1993).
- ⁵G. A. Athanassoulis and A. M. Prospathopoulos, "3D acoustic scattering of source-generated acoustic field by an axisymmetric island," in *Proceedings of the XIth Symposium on Hydroacoustics*, Gdynia-Jurata, Poland, 24–27 May 1994, pp. 75–82.
- ⁶G. A. Athanassoulis and A. M. Prospathopoulos, "Three-dimensional acoustic scattering of a source generated field from a cylindrical island," *J. Acoust. Soc. Am.* **100**, 206–218 (1996).
- ⁷J. R. Higgins, *Completeness and Basis Properties of Sets of Special Functions* (Cambridge U.P., Cambridge, 1977).
- ⁸C. A. Boyles, *Acoustic Waveguides: Applications to Ocean Science* (Wiley, New York, 1984).
- ⁹G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (Prentice-Hall, London, 1994).
- ¹⁰F. B. Jensen, W. A. Kuperman, M. A. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, Woodbury, NY, 1994).
- ¹¹G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1966).
- ¹²M. A. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (Dover New York, 1972), 9th ed.
- ¹³I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
- ¹⁴F. Ursell, "Creeping modes in a shadow," *Proc. Camb. Philos. Soc.* **64**, 171–191 (1968).
- ¹⁵R. Burridge and H. Weinberg, "Horizontal rays and vertical modes," in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, Berlin, 1977), pp. 86–152.
- ¹⁶P. Henrici, *Applied and Computational Complex Analysis* (Wiley, New York, 1977), Vol. II.
- ¹⁷Y. V. Sidorov, M. V. Fedoryuk, and M. I. Shaburin, *Lectures on the Theory of Functions of a Complex Variable* (MIR, Moscow, 1985).
- ¹⁸M. J. Buckingham, "Stationary phase evaluation of the integral for the acoustic field around a conical seamount," *J. Acoust. Soc. Am.* **80**, 278–281 (1986).

Observations on attenuation and shear-wave velocity in fine-grained, marine sediments

Frederick A. Bowles

Naval Research Laboratory, Marine Geosciences Division, Stennis Space Center, Mississippi 39529

(Received 29 July 1996; accepted for publication 16 December 1996)

Compressional(*P*)-wave attenuation, shear(*S*)-wave attenuation, and shear(*S*)-wave velocity measurements, compiled for fine-grained, unconsolidated sediments, show trends that do not support empirical relationships [Hamilton, *J. Acoust. Soc. Am.* **68**, 1313–1340 (1980); in *Acoustics and Ocean Bottom* (F. A. S. E. Specialize Conf., Madrid, 1987)] commonly used for geoacoustic modeling. The exception is compressional-wave attenuation data, from 10 Hz to 500 kHz, that essentially follow a frequency dependence of f^1 (± 2 standard deviations). *P*-wave attenuation with depth is difficult, if not impossible, to predict given the variability in the data. Examination of several attenuation/depth profiles indicates that those of Mitchell and Focke [*J. Acoust. Soc. Am.* **67**, 1582–1589 (1980)] may be the best choice for predicting *P*-wave attenuation with depth at the present time. The assumption that *S*-wave and *P*-wave attenuations are proportional [Hamilton, *J. Acoust. Soc. Am.* **60**, 334–338 (1976c); in *Acoustics and Ocean Bottom* (F. A. S. E. Specialize Conf., Madrid, 1987)] is not supported by the data. *S*-wave attenuations calculated from effective stress mimick the data, however, suggesting this as a better method for predicting *S*-wave attenuation with depth. *S*-wave velocity/depth profiles fall into high- and low-velocity groups that (1) reflect gross differences in sediment texture, and (2) demonstrate the importance of sediment characterization in order to select an appropriate velocity/depth function. [S0001-4966(97)06505-3]

PACS numbers: 43.30.Es, 43.30.Ma [JHM]

INTRODUCTION

The importance of geoacoustic models in support of acoustic experiments has long been stressed by Hamilton (1974, 1980, 1987). Indeed, his work is widely used by the acoustic community today for predicting sediment geoacoustic properties in lieu of direct measurements. These properties are typically furnished to acousticians as range-independent models that incorporate the variation of each geoacoustic property as a function of depth. In constructing such models, estimates of compressional (*P*)-wave and shear (*S*)-wave attenuation and shear (*S*)-wave velocity are particularly suspect because they cannot be easily or accurately measured.

The latest compilations of *P*-wave attenuation were published in 1987 and 1989 by Hamilton and Kibblewhite, respectively. Also, a significant body of *S*-wave velocity information in the literature has never been assembled for comparison or evaluation of its usefulness to geoacoustic modeling. Hence, the objectives of this paper are to (1) provide updated compilations of *P*-wave attenuation, *S*-wave attenuation, and *S*-wave velocity data, (2) present the compiled data efficiently (tables and graphs) for easy assessment and use by modelers to better predict these parameters, and (3) evaluate the utility of Hamilton's empirical relationships for predicting the above parameters in light of more recent data. For the purposes of this paper, all the data are assumed to be equally reliable.

I. APPROACH

This work originates from a more comprehensive review of published geoacoustic properties data (Bowles, 1994) in

order to develop a generic geoacoustic model relating to calcareous clay sediment, and to support acoustic experiments at selected sites located east of Bermuda on the lower flanks of the Mid-Atlantic Ridge (Caruthers *et al.*, 1991; Tucholke *et al.*, 1991). The compiled data were limited to mostly fine-grained sediments because (1) deposition in the region is predominantly pelagic, and (2) there are basic differences between the responses of fluid-saturated, coarse-grained, and fine-grained sediments (Stoll, 1985).

Compressional-wave attenuations for fine-grained sediments (Table I) were culled from the compilations of Hamilton (1972, 1976a, 1987). No additional fine-grained sediment data resulted from Kibblewhite's review (1989), but a search of subsequent literature uncovered several relevant publications (Chotiros, 1989; Richardson *et al.*, 1992; Courtney and Mayer, 1993; Rogers *et al.*, 1993; and Tang *et al.*, 1994). In the case of *S*-wave attenuation, there are no compilations upon which to build. Furthermore, there are relatively few investigations of *S*-wave attenuation and some of these involve sandy sediments (Table II).

Tables I and II contain sediment types such as sand-silt-clay, mud, and turbidite. These descriptions are ambiguous because they do not identify whether the major textural component of the sediment is fine or coarse grained. In most of these cases the attenuation values were indicative of a fine-grained character; where this was clearly not the case, the data were not used. Other than sediment texture, no other factor was used to "filter" the data.

These compilations do not include samples below 300 m sub-bottom. This was considered (Bowles, 1994) to be a reasonable thickness for ponded sediments within the ridge

TABLE I. Compressional-wave attenuation data.

| Reference | Publication Year | Sediment type | Depth (m) | Frequency (Hz) | Attenuation (dB/m) | K (dB/m/kHz) |
|--------------------------|---------------------|-------------------------|--------------|-------------------|-----------------------|-------------------|
| Shumway | 1960 | clayey silt (ave. 12) | 0 | 20 500 | 1.09 | 0.053 |
| Shumway | 1960 | clayey silt (ave. 16) | 0 | 22 800 | 1.67 | 0.0732 |
| Shumway | 1960 | clayey silt (ave. 20) | 0 | 30 400 | 3.2 | 0.105 |
| Wood and Weston | 1964 | mud | 0 | 4000 | 0.279 | 0.0698 |
| Wood and Weston | 1964 | mud | 0 | 8000 | 0.476 | 0.0595 |
| Wood and Weston | 1964 | mud | 0 | 16 000 | 0.951 | 0.0594 |
| Wood and Weston | 1964 | mud | 0 | 32 000 | 1.8 | 0.0564 |
| Wood and Weston | 1964 | mud | 0 | 48 000 | 3.02 | 0.0663 |
| Bennett | 1966 | abyssal plains (ave) | 0 | 12 000 | 0.7 | 0.058 |
| Ulonka | 1968 | clayey silt | 0 | 250 000 | 26.2 | 0.105 |
| McLeroy and DeLoach | 1968 | silty clay (site 6) | 0 | 15 000 | 1.08 | 0.072 |
| McLeroy and DeLoach | 1968 | clay-silt (site 9) | 0 | 15 000 | 1.12 | 0.075 |
| McCann and McCann | 1969 | clay | 0 | 368 000 | 31.9 | 0.087 |
| Berzon <i>et al.</i> | 1969 | clay | 140 | 25 | 0.0035 | 0.139 |
| Anderson and Blackman | 1971 | calcareous clay | 250 | 200 | 0.0031 | 0.016 |
| Anderson and Blackman | 1971 | calcareous clay | 250 | 315 | 0.0126 | 0.04 |
| Anderson and Blackman | 1971 | calcareous clay | 250 | 250 | 0.0069 | 0.028 |
| Neprochnov | 1971 | layer 1, Bengal Bay | 350 | 110 | 0.0165 | 0.15 |
| Neprochnov | 1971 | layer 1, Arabian Sea | 400 | 110 | 0.0156 | 0.142 |
| Neprochnov | 1971 | layer 1, S. Indian Ocn | 200 | 110 | 0.02 | 0.182 |
| Neprochnov | 1971 | layer 1, N. Indian Ocn | 400 | 110 | 0.0182 | 0.166 |
| Neprochnov | 1971 | layer 1, Japan Sea | 250 | 110 | 0.02 | 0.182 |
| Neprochnov | 1971 | layer 1, Black Sea | 350 | 110 | 0.0165 | 0.15 |
| Neprochnov | 1971 | Ob Trench | 350 | 110 | 0.0174 | 0.158 |
| Schirmer | 1971 | mud (clay-silt) | 0 | 1200 | 0.29 | 0.242 |
| Hamilton | 1972 | clayey silt | 0 | 7000 | 1.2 | 0.171 |
| Hamilton | 1972 | clayey silt | 0 | 14 000 | 2.4 | 0.171 |
| Hamilton | 1972 | clayey silt | 0 | 14 000 | 2.3 | 0.164 |
| Hamilton | 1972 | clayey silt | 0 | 14 000 | 1 | 0.071 |
| Hamilton | 1972 | clayey silt | 0 | 14 000 | 0.6 | 0.043 |
| Hamilton | 1972 | clayey silt | 0 | 23 200 | 2.8 | 0.121 |
| Hamilton | 1972 | clayey silt | 0 | 25 000 | 4 | 0.16 |
| Hamilton | 1972 | clayey silt | 0 | 100 000 | 18 | 0.18 |
| Tyce | 1976 | carbonate | 28 | 4000 | 0.12 | 0.03 |
| Dicus | 1976 | turbidite | 106 | 65 | 0.002 | 0.031 |
| Helmberger <i>et al.</i> | 1979 | ? (Bering Sea) | 150 | 10 | 0.000 295 | 0.0295 |
| Ferla <i>et al.</i> | 1980 | mud | 0 | 100 | 0.0041 | 0.041 |
| Ferla <i>et al.</i> | 1980 | mud | 1 | 250 | 0.017 | 0.068 |
| Wroldstad | 1980 | turbidite | 228 | 80 | 0.004 | 0.05 |
| Chapman | 1980 | turbidite | 200 | 160 | 0.002 | 0.013 |
| Mitchell and Focke | 1980 | turbidite | 0 | 213 | 0.0085 | 0.04 |
| Mitchell and Focke | 1980 | turbidite | 0 | 213 | 0.0055 | 0.026 |
| Rubano | 1980 | clay/silt | 0 | 20 | 0.0006 | 0.03 |
| Rubano | 1980 | clay/silt | 0 | 250 | 0.0094 | 0.038 |
| Hooper <i>et al.</i> | 1981 | turbidite | 0 | 163 | 0.0054 | 0.086 |
| Frisk <i>et al.</i> | 1981 | silt/clay | 80 | 220 | 0.0015 | 0.007 |
| Tyce | 1981 | clayey silt? | 14 | 4000 | 0.29 | 0.072 |
| Tyce | 1981 | clayey silt? | 16 | 4000 | 0.25 | 0.062 |
| Tyce | 1981 | clayey silt? | 18 | 4000 | 0.29 | 0.072 |
| Tyce | 1981 | hemipelagic turbid. | 27 | 4000 | 0.26 | 0.065 |
| Tyce | 1981 | hemipelagic turbid. | 21 | 4000 | 0.19 | 0.048 |
| Tyce | 1981 | hemipelagic turbid. | 25 | 4000 | 0.28 | 0.07 |
| Tyce | 1981 | hemipelagic clay | 14 | 4000 | 0.26 | 0.065 |
| Tyce | 1981 | clayey silt? | 15 | 4000 | 0.21 | 0.052 |
| Dicus and Anderson | 1982 | turbidite | 0 | 65 | 0.0026 | 0.04 |
| Dicus and Anderson | 1982 | turbidite | 0 | 65 | 0.0033 | 0.051 |
| Dicus and Anderson | 1982 | calcareous clay | 0 | 65 | 0.0176 | 0.271 |
| Dicus and Anderson | 1982 | turbidite | 0 | 65 | 0.0013 | 0.02 |
| Tindle | 1982 | silt/clay | 50 | 80 | 0.0045 | 0.056 |
| Tindle | 1982 | silt/clay | 50 | 140 | 0.0078 | 0.056 |
| Stoll and Houtz | 1983 | sand/silt/clay | 350 | 60 | 0.007 56 | 0.126 |
| Stoll and Houtz | 1983 | sand/silt/clay | 350 | 10 | 0.000 13 | 0.013 |
| Stoll and Houtz | 1983 | sand/silt/clay | 350 | 20 | 0.001 13 | 0.056 |
| Focke | 1983 | silt/clay | 0 | 200 | 0.006 | 0.027 |
| Jacobson <i>et al.</i> | 1984 | turbidite | 0 | 95 | 0.004 | 0.042 |
| Barker and Helmberger | 1986 | ? (Bering Sea) | 150 | 32 | 0.0014 | 0.045 |
| Barker and Helmberger | 1986 | ? (Bering Sea) | 150 | 18 | 0.0005 | 0.027 |
| Frisk <i>et al.</i> | 1986 | silt/clay | 26 | 220 | 0.0039 | 0.018 |
| Frisk <i>et al.</i> | 1986 | silt/clay | 19 | 220 | 0.007 | 0.032 |
| Duckworth and Baggeroer | 1986 | turbidite | 100 | 27.5 | 0.0019 | 0.069 |
| Chotiros | 1989 | clay (ave. 26) | 30 | 400 000 | 39.1 | 0.098 |
| Richardson <i>et al.</i> | 1992 | Holocene clays | 0 | 400 000 | 59 | 0.15 |
| Richardson <i>et al.</i> | 1992 | Hemipelagic silty clays | 0 | 400 000 | 50 | 0.12 |
| Courtney & Mayer | 1993 | LaHave clay | 2.5 | 500 000 | 70 | 0.14 |
| Courtney & Mayer | 1993 | LaHave clay | 6 | 500 000 | 90 | 0.18 |
| Rogers <i>et al.</i> | 1993 | sand and silty clay | 100 | 50 | 0.000 557 | 0.0111 |
| Rogers <i>et al.</i> | 1993 | sand and silty clay | 100 | 75 | 0.000 753 | 0.01 |
| Rogers <i>et al.</i> | 1993 | sand and silty clay | 100 | 175 | 0.002 13 | 0.0123 |
| Rogers <i>et al.</i> | 1993 | sand and silty clay | 100 | 275 | 0.003 61 | 0.0131 |
| Rogers <i>et al.</i> | 1993 | sand and silty clay | 100 | 375 | 0.005 24 | 0.014 |
| Rogers <i>et al.</i> | 1993 | sand and silty clay | 100 | 525 | 0.008 02 | 0.0153 |
| Rogers <i>et al.</i> | 1993 | sand and silty clay | 100 | 600 | 0.009 55 | 0.0159 |
| Tang <i>et al.</i> | 1994 | silty clay | 0 | 58000 | 4.11 | 0.0709 |

TABLE II. Shear-wave attenuation data.

| Reference | Publication year | Sediment type | Depth (m) | Frequency (Hz) | Shear vol. (m/s) | Log decrement (dB/wavelength) | Attenuation (dB/m) | K (dB/m/kHz) |
|--------------------------------|---------------------|-------------------|--------------|-------------------|---------------------|----------------------------------|-----------------------|-------------------|
| Bucker <i>et al.</i> | 1964 | sand | 0 | 20 | 101 | 1.66 | 0.328 | 16.4 |
| Bucker <i>et al.</i> | 1964 | sand | 0 | 25 | 197 | 1.55 | 0.197 | 7.88 |
| Warrick | 1974 | bay mud | 1 | 5 | 90 | 1.71 | 0.095 | 18.9 |
| Joyner <i>et al.</i> | 1976 | bay mud | 0 | 5 | 30 | 1.71 | 0.284 | 56.8 |
| Schirmer | 1980 | sand | 3 | 4.5 | 111 | 0.173 | 0.007 | 1.56 |
| Holt <i>et al.</i> | 1983 | sand | 0 | 35 | 135 | 2.3 | 0.6 | 17.1 |
| Jensen and Schmidt | 1986 | sand/silt | 0 | 3 | 85 | 0.45 | 0.016 | 5.29 |
| Jensen and Schmidt | 1986 | sand/silt | 60 | 3 | 300 | 0.15 | 0.002 | 0.5 |
| Rauch | 1986 | soft mud | 2 | 3.5 | 125 | 0.2 | 0.0056 | 1.6 |
| Carter <i>et al.</i> | 1986 | biosiliceous clay | 178 | 18 | 240 | 0.273 | 0.0205 | 1.137 |
| Sauter | 1987 | clay/ooze | 0 | 4 | 30 | 1.229 | 0.164 | 40.97 |
| Sauter | 1987 | clay/ooze | 0.5 | 4 | 32 | 1.092 | 0.136 | 34.11 |
| Sauter | 1987 | clay/ooze | 1 | 4 | 34 | 0.954 | 0.112 | 28.06 |
| Sauter | 1987 | clay/ooze | 2 | 4 | 36 | 0.819 | 0.091 | 22.76 |
| Sauter | 1987 | clay/ooze | 2.5 | 4 | 38 | 0.682 | 0.072 | 17.95 |
| Sauter | 1987 | clay/ooze | 3 | 4 | 40 | 0.546 | 0.055 | 13.64 |
| Sauter | 1987 | clay/ooze | 3.5 | 4 | 42 | 0.409 | 0.039 | 9.74 |
| Sauter | 1987 | clay/ooze | 6 | 4 | 54 | 0.273 | 0.02 | 5.05 |
| Sauter | 1987 | clay/ooze | 8 | 4 | 63 | 0.226 | 0.014 | 3.59 |
| Sauter | 1987 | clay/ooze | 10 | 4 | 70 | 0.191 | 0.011 | 2.73 |
| Sauter | 1987 | clay/ooze | 12 | 4 | 73 | 0.177 | 0.01 | 2.43 |
| Sauter | 1987 | clay/ooze | 14 | 4 | 77 | 0.164 | 0.0085 | 2.13 |
| Sauter | 1987 | clay/ooze | 16 | 4 | 79 | 0.15 | 0.0076 | 1.9 |
| Sauter | 1987 | clay/ooze | 18 | 4 | 82 | 0.145 | 0.0071 | 1.76 |
| Sauter | 1987 | clay/ooze | 20 | 4 | 85 | 0.136 | 0.0064 | 1.61 |
| Sauter | 1987 | clay/ooze | 40 | 4 | 140 | 0.115 | 0.0033 | 0.82 |
| Sauter | 1987 | clay/ooze | 60 | 4 | 180 | 0.102 | 0.0023 | 0.57 |
| Sereno | 1990 | zeolitic Clay | 35 | 15 | 116 | 0.682 | 0.088 | 5.88 |
| Caiti <i>et al.</i> | 1991 | mud | 0 | 8 | 68 | 1.2 | 0.14 | 17.6 |
| Caiti <i>et al.</i> | 1991 | mud | 1 | 8 | 74 | 1.1 | 0.12 | 14.9 |
| Caiti <i>et al.</i> | 1991 | mud | 2 | 8 | 83 | 1 | 0.1 | 12 |
| Caiti <i>et al.</i> | 1991 | mud | 3 | 8 | 97 | 0.9 | 0.07 | 9.3 |
| Caiti <i>et al.</i> | 1991 | mud | 4 | 8 | 115 | 0.7 | 0.05 | 6.1 |
| Caiti <i>et al.</i> | 1991 | mud | 7 | 8 | 121 | 0.6 | 0.04 | 5 |
| Caiti <i>et al.</i> | 1991 | mud | 15 | 8 | 139 | 0.5 | 0.03 | 3.6 |
| Muir <i>et al.</i> | 1991 | mud | 0 | 4 | 45 | 0.475 | 0.042 | 10.55 |
| Muir <i>et al.</i> | 1991 | mud | 2.5 | 4 | 90 | 0.38 | 0.017 | 4.22 |
| Muir <i>et al.</i> | 1991 | mud | 5 | 4 | 103 | 0.31 | 0.012 | 3.01 |
| Muir <i>et al.</i> | 1991 | mud | 7.5 | 4 | 125 | 0.26 | 0.008 | 2.08 |
| Muir <i>et al.</i> | 1991 | mud | 15 | 4 | 134 | 0.18 | 0.0054 | 1.34 |
| Muir <i>et al.</i> | 1991 | mud | 27.5 | 4 | 141 | 0.165 | 0.0047 | 1.17 |
| Muir <i>et al.</i> | 1991 | mud | 0 | 20 | 55 | 0.6 | 0.218 | 10.9 |
| Muir <i>et al.</i> | 1991 | mud | 1 | 20 | 68 | 0.25 | 0.074 | 3.68 |
| Muir <i>et al.</i> | 1991 | mud | 2 | 20 | 73 | 0.25 | 0.068 | 3.42 |
| Muir <i>et al.</i> | 1991 | mud | 3 | 20 | 95 | 0.4 | 0.084 | 4.21 |
| Muir <i>et al.</i> | 1991 | mud | 4 | 20 | 112 | 0.25 | 0.045 | 2.23 |
| Muir <i>et al.</i> | 1991 | mud | 0 | 30 | 55 | 1 | 0.545 | 18.2 |
| Muir <i>et al.</i> | 1991 | mud | 1 | 30 | 68 | 0.65 | 0.287 | 9.56 |
| Muir <i>et al.</i> | 1991 | mud | 2 | 30 | 73 | 0.5 | 0.205 | 6.85 |
| Muir <i>et al.</i> | 1991 | mud | 3 | 30 | 95 | 0.75 | 0.237 | 7.89 |
| Muir <i>et al.</i> | 1991 | mud | 4 | 30 | 112 | 0.5 | 0.134 | 4.46 |
| Muir <i>et al.</i> | 1991 | mud | 0 | 40 | 55 | 1.5 | 1.09 | 27.3 |
| Muir <i>et al.</i> | 1991 | mud | 1 | 40 | 68 | 1.25 | 0.735 | 18.4 |
| Muir <i>et al.</i> | 1991 | mud | 2 | 40 | 73 | 0.6 | 0.329 | 8.22 |
| Muir <i>et al.</i> | 1991 | mud | 3 | 40 | 95 | 0.8 | 0.337 | 8.42 |
| Muir <i>et al.</i> | 1991 | mud | 4 | 40 | 112 | 0.6 | 0.214 | 5.36 |
| Muir <i>et al.</i> | 1991 | mud | 0 | 80 | 55 | 4.6 | 6.69 | 83.6 |
| Muir <i>et al.</i> | 1991 | mud | 1 | 80 | 68 | 2.6 | 3.06 | 38.2 |
| Muir <i>et al.</i> | 1991 | mud | 2 | 80 | 73 | 1.9 | 2.08 | 26 |
| Muir <i>et al.</i> | 1991 | mud | 3 | 80 | 95 | 1.9 | 1.6 | 20 |
| Muir <i>et al.</i> | 1991 | mud | 4 | 80 | 112 | 2.15 | 1.54 | 19.2 |
| Muir <i>et al.</i> | 1991 | mud | 0 | 100 | 55 | 6.7 | 12.2 | 121.8 |
| Muir <i>et al.</i> | 1991 | mud | 1 | 100 | 68 | 4.9 | 7.2 | 72 |
| Muir <i>et al.</i> | 1991 | mud | 2 | 100 | 73 | 3.2 | 4.38 | 43.8 |
| Muir <i>et al.</i> | 1991 | mud | 3 | 100 | 95 | 2.3 | 2.42 | 24.2 |
| Muir <i>et al.</i> | 1991 | mud | 4 | 100 | 112 | 2.75 | 2.46 | 24.6 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 0 | 20 | 110 | 0.62 | 0.113 | 5.64 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 1 | 20 | 170 | 0.568 | 0.067 | 3.34 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 2 | 20 | 214 | 0.395 | 0.037 | 1.85 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 4 | 20 | 225 | 0.246 | 0.022 | 1.09 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 7 | 20 | 254 | 0.172 | 0.014 | 0.676 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 10 | 20 | 276 | 0.159 | 0.012 | 0.575 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 15 | 20 | 297 | 0.159 | 0.011 | 0.531 |
| Stoll <i>et al.</i> | 1991 | clayey silt | 20 | 20 | 310 | 0.159 | 0.01 | 0.512 |
| Bibee and Dorman | 1991 | hemipelagic clay | 5 | 2.5 | 40 | 0.758 | 0.05 | 18.95 |
| Bibee and Dorman | 1991 | hemipelagic clay | 5 | 2.5 | 40 | 0.303 | 0.019 | 7.6 |
| Brominski <i>et al.</i> | 1992 | biosiliceous clay | 180 | 18 | 200 | 0.281 | 0.0025 | 1.4 |
| Brominski <i>et al.</i> | 1992 | biosiliceous clay | 180 | 3 | 200 | 0.281 | 0.0042 | 1.4 |
| Liu <i>et al.</i> | 1994 | hemipelagic clay | 2.5 | 1.5 | 40 | 0.546 | 0.02 | 13.6 |
| Ali and Broadhead | 1995 | bay mud | 2 | 54 | 122 | 1.71 | 0.755 | 14 |
| Bibee (personal communication) | 1995 | high porosity mud | 0 | 2 | 12 | 2.2 | 0.367 | 183 |
| Bibee (personal communication) | 1995 | high porosity mud | 0.5 | 2 | 14 | 1.58 | 0.219 | 110 |
| Bibee (personal communication) | 1995 | high porosity mud | 1 | 2 | 18 | 1.07 | 0.117 | 58 |
| Bibee (personal communication) | 1995 | high porosity mud | 1.5 | 2 | 22 | 0.812 | 0.073 | 37 |
| Bibee (personal communication) | 1995 | high porosity mud | 2 | 2 | 25 | 0.68 | 0.054 | 27 |
| Bibee (personal communication) | 1995 | high porosity mud | 2.5 | 2 | 29 | 0.62 | 0.044 | 22 |

flank topography west of the Bermuda pedestal (fracture zones excluded).

Shear-wave velocities based on Deep Sea Drilling Project and Ocean Drilling Program (DSDP/ODP) measurements were compiled and plotted against measurement depth. Only the linear curve fit of the data (Fig. 8 in Bowles, 1994) is presented here, along with other velocity-depth profiles extracted from the literature. Most of the additional profiles are smoothed versions of the original data.

II. OBSERVATIONS

A. Compressional-wave attenuation

P -wave attenuation measurements (Table I) plotted against frequency (Fig. 1) cluster in three areas with significant data gaps between 600 Hz–4 kHz and 60–250 kHz. Although there appears to be no fundamental reason for these gaps, there are practical considerations. The low-frequency gap probably reflects the trade-off between acoustic penetration and acoustic resolution that determines, in turn, the frequency of existing seismic (field) instrumentation. The high-frequency gap, on the other hand, marks a transitional region where frequencies used for conventional field measurements are less suitable (or not at all) for measurements made in the laboratory, and vice versa. The data from about 15 to 30 kHz consist of both laboratory and field measurements. This has considerable significance because both measurement techniques, although different in approach and scale, resulted in attenuations of similar magnitude.

The data from 10 to 600 Hz are widely scattered. The spread most likely reflects differences in sediment type, degree of consolidation, diagenesis, layering, as well as other causes of effective attenuation encountered during low-frequency field measurements. Moreover, attenuation varies with depth, and longer wavelengths (i.e., low frequencies) encounter deeper sediments with unknown properties.

A curve fit (heavy line) of the data in Fig. 1 reveals that a best fit is achieved when attenuation (α) is dependent on frequency (f) to the power of ~ 1.12 , and may be expressed by

$$\alpha(\text{dB/m}) = 2.42 \times 10^{-5} f^{1.12}.$$

For the most part, the acoustic community uses empirical relationships developed by Hamilton to estimate P -wave attenuation (Kibblewhite, 1989). Hamilton's view, most recently stated in 1987, is that P -wave attenuation varies with the first power of frequency (f^1), the tacit assumption being that anelastic frame losses dominate in producing an attenuation. Alternatively, the Biot–Stoll model (Stoll and Bryan, 1970) is based on the premise that attenuation associated with the movement of pore fluids modifies the normal f^1 relationship between frequency and attenuation. Hence, applying Hamilton's simple linear viscoelastic model to predict attenuations from seismic to ultrasonic frequencies is unjustified. Both perspectives are represented in Fig. 1: Stoll's, by the stippling, which represents the range of predicted attenuation values for low-permeability sediments (i.e., silts, clays) calculated from the Biot–Stoll model; and Hamilton's, by the f^1 line, which represents the slope of a line with a first-

power frequency dependence (for comparison, it is shown passing through the midpoint of the $f^{1.12}$ line). Agreement between the stippling and $f^{1.12}$ line is extremely good. On the other hand, the slopes of the f^1 and $f^{1.12}$ lines are not substantially different. Indeed, a t test of the data shows that the difference is insignificant at 95% confidence limits (two standard deviations). It is not possible, therefore, to differentiate among f^1 , $f^{1.12}$, or the Biot–Stoll model given the variability of the data. Therefore, use of a linear relationship between attenuation and frequency is acceptable from 10 Hz to 500 kHz for predicting attenuations in fine-grained (low permeability), unconsolidated sediments.

Profiles (Fig. 2) based on data compiled by Hamilton (1976a) are commonly used to predict P -wave attenuation with depth. The attenuations, it will be noted, are expressed as the coefficient K . Hence, to be consistent with Hamilton, the new data are also presented in this form. The silt-clay profile (Fig. 2) is relevant to this study, but is obviously a visual-fit of a few, scattered data points. When updated with newer data, the fit (profile 1 in Fig. 3) is no better. The data, in fact, are too scattered to warrant any curve fit (the scatter is not surprising considering the widely varying field conditions represented by these measurements). Nevertheless, profiles from Mitchell and Focke (1980), Jacobson *et al.* (1981), and Brienzo (1992) are included in Fig. 3 for comparison and because they are the choices that modelers presently have for predicting P -wave attenuation with sediment depth. Given these choices, which profile provides the best prediction of

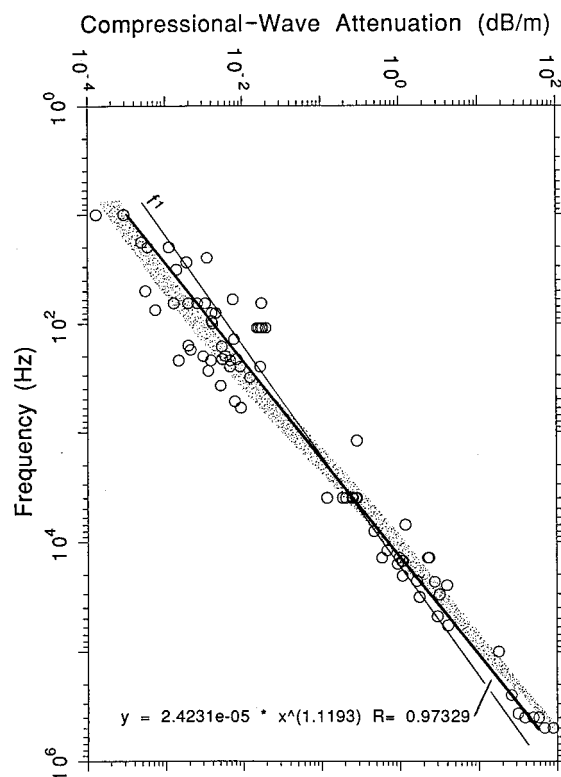


FIG. 1. Compressional-wave attenuation versus frequency for fine-grained marine sediments (circles represent the data listed in Table I). The heavy line represents a computer-generated power curve fit of the data and the line labeled f^1 represents the slope of a hypothetical data set having a dependence of attenuation on the first power of frequency. The stippled area defines the range for attenuation values for silts and clays calculated from the Biot–Stoll model (from Stoll, 1989; see Fig. 5.13).

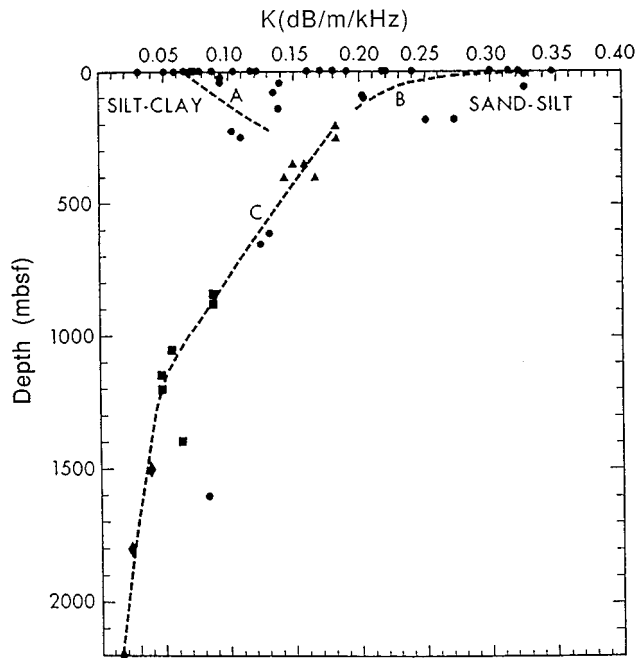


FIG. 2. Measurements (circles, squares, triangles) of compressional-wave attenuation (expressed as the attenuation coefficient k) versus depth in the seafloor (from Hamilton, 1976a).

attenuation with depth in areas with predominantly fine-grained sediments?

A review of the data used to prepare Fig. 2 indicates that Hamilton's silt-clay profile is influenced by measurements that were made in relatively sandy sediments. Indeed, if the new data that are plotted in Fig. 3 were also plotted in Fig. 2, the points would fall well to the left of his silt-clay profile. Thus, Hamilton's profile predicts attenuation values that are generally too high for purely fine-grained deposits.

Jacobson *et al.*'s (1981) profile, which coincides approximately with Hamilton's (Fig. 3), is derived from seismic refraction data collected in the northern Bay of Bengal. The region consists of fine-grained sediments interbedded with thick, sand/silt turbidites. The attenuations, then, are influenced not only by coarse-grained sediments, but also by bedding planes that enhance scattering attenuation. Thus, use of Jacobson's profile for modeling fine-grained sediments is also questionable.

Brienzo (1992) used spectral ratios to generate estimates of P -wave attenuation for the Monterey Fan, also an area of turbidite deposition. His profile (4), however, reflects attenuation values that are significantly lower than those reported by Jacobson *et al.* (1981). The experimental site is located on a relatively flat portion of the upper fan near Monterey canyon/channel, which funnels coarse sediment downslope. The low attenuation values reflect the silty clay character of the sediments that are deposited in the off-channel areas of the fan (Brienzo, 1992).

The profiles of Mitchell and Focke (1980) derive from bottom-loss data collected in regions of predominantly calcareous ooze-calcareous clay deposition and silty clay deposition (Fig. 3, profiles 5 and 6, respectively). Jacobson *et al.* (1981) have noted that these profiles retain essentially the

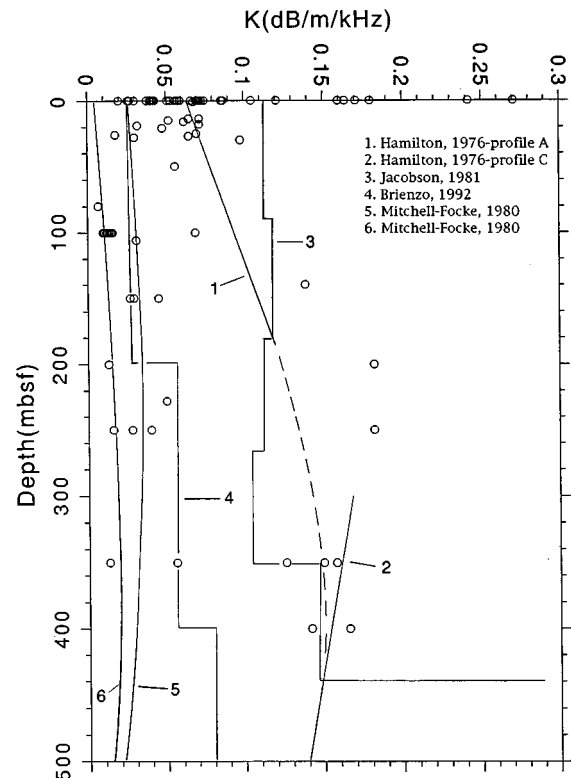


FIG. 3. Compressional-wave attenuation (expressed as the attenuation coefficient k) versus depth in the seafloor for fine-grained sediments (circles represent the data listed in Table I). Solid lines represent attenuation/depth profiles based on specific field experiments (as indexed according to the figure legend). Profiles 1 and 2 (heavy lines) represent the A and C profiles, respectively, shown in Fig. 2.

same trend as Hamilton's profiles (1 and 2 in Fig. 3), but that the attenuation values are lower by an order of magnitude.

Given the apparent coarse-grained bias of Hamilton's profile, either of the Mitchell–Focke profiles is probably a better choice for predicting P -wave attenuation with depth in fine-grained sediments. Several works support this opinion. Matthews (1982) observed that Mitchell–Focke attenuation values yield plane-wave, bottom-loss calculations that agree closely with measured bottom-loss values. Hampton (1985) showed that Mitchell–Focke attenuations compare favorably with Biot theory. He also noted that low values of sediment attenuation (*vis-à-vis*, high values) have successfully explained acoustic propagation effects. Rogers *et al.* (1993) reported that their attenuation values for the upper 100 m of sandy/silty clays agree well with Mitchell and Focke values. Finally, it is apparent (Fig. 3) that Brienzo's profile compares favorably with the Mitchell–Focke profiles, particularly in the first 200 meters.

The following regression equations will essentially reproduce the Mitchell–Focke profiles for the upper 500 m of sediment:

$$K_H = 0.0262 + 8.301 \times 10^{-5}d + (-2.483 \times 10^{-7})d^2 + 1.455 \times 10^{-7}d^3,$$

$$K_L = 0.0044 + 4.294 \times 10^{-5}d + 1.974 \times 10^{-7}d^2 + (-7.948 \times 10^{-10})d^3 + 6.342 \times 10^{-13}d^4,$$

where K_H refers to curve 5, K_L refers to curve 6, d =depth, and K is in dB/m/kHz (Fig. 3).

Limiting the above discussion to sediment type (i.e., grain size) is not meant to imply that attenuations associated with the mineral frame and pore-fluid motions principally determine the nature of the profiles (Fig. 3). The limitation is largely due to a lack of related information. The profiles, in fact, reflect the summation of all attenuations (effective attenuation) which includes losses associated with sediment layering (interbed multiples), energy conversion (between compression, shear, and interface waves), and volume inhomogeneities (e.g., gas bubbles).

B. Shear-wave attenuation

Hamilton (1980, 1987) assumes that S -wave attenuation varies with depth in the seafloor proportionally with P -wave attenuation, according to the following equation:

$$K_{S_z} = (K_{S_0}/k_{p0})K_{p_z},$$

where K_{S_z} is shear-wave attenuation at some depth (z), K_{S_0} is shear-wave attenuation at zero depth (i.e., the seafloor), K_{p0} is compressional-wave attenuation at zero depth, and K_{p_z} is compressional-wave attenuation at depth (z). Thus, in this model, S -wave attenuation/depth profiles are expected to track P -wave attenuation/depth profiles (Fig. 2), but at substantially higher K values. Hamilton (1976c), for example, suggests using $K_{S_0} = 17.3$ dB/m/kHz as an initial (surface) value for unconsolidated clays, whereas K_{p0} values are in the 0.02–0.27 dB/m/kHz range.

Published S -wave attenuation data (Table II) are plotted against depth in Fig. 4, but in dB/ λ (i.e., log decrement). The reason for using log decrement (δ) instead of the coefficient K , as before (Figs. 2 and 3), is because most of the shear-wave data are reported either in this form or as the quality factor Q , which readily converts to log decrement by the formula $\delta(\text{dB}/\lambda) = 8.686\pi/Q$. Although K values are provided (Table II), converting to K is not as straightforward as converting to δ and leaves room for error.

The attenuation measurements (Fig. 4) are confined mostly to the upper 5 m of sediment where they range in value between 0.1–6.7 dB/ λ . Below 5 m the values fall between about 0.1–0.8 dB/ λ with most of them less than 0.3 dB/ λ . Data points greater than 1.8 dB/ λ are largely due to 80 and 100 Hz measurements (Table II). Most of the data are concentrated in the upper 20 m of sediment (Fig. 5; the 80 and 100 Hz data are omitted) and represent primarily Scholte-wave or short-borehole measurements (profiles 4, 5, and 6). All the profiles in Fig. 5 show attenuation decreasing, often rapidly, within the upper 4–5 m of sediment. For the omitted frequencies (80 and 100 Hz), the decrease is exceedingly sharp and occurs mostly in the upper 2–3 m of sediment (Fig. 4). Interestingly, profiles 1, 3, and 7 are nearly coincident below about 10 m, as would profile 4 if it extended beyond 4 m (were it not for the attenuation “spike” at 3 m, the 30 and 40 Hz profiles would probably coincide as well). Sauter’s work (profile 1) actually extends to 60 m (Fig. 4) where it coincides with a value reported by Jensen and Schmidt (1986). Furthermore, when profiles 1, 3, and 7 are extrapolated to greater depths they coincide (nearly) with

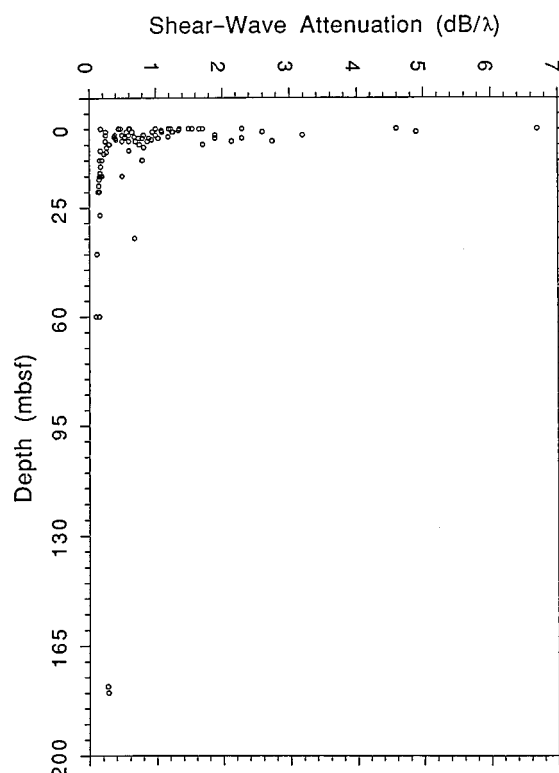


FIG. 4. Shear-wave attenuation (expressed as log decrement) versus depth in the seafloor for fine-grained sediments (circles represent the data listed in Table II).

attenuations reported by Carter *et al.* (1986) and Bromirski *et al.* (1992), also plotted in Fig. 4 at 180 m (midpoint of the sediment section in which the measurements were made).

Given the small amount of data, measurement difficulties, and inherent site differences, it is still notable that the trends exhibited by the individual profiles in Fig. 5, as well as the total data (Fig. 4), are consistent with theoretical considerations (Stoll, 1989) predicting large attenuations at the seafloor, rapidly decreasing attenuations in the upper few meters, and nearly constant, low values thereafter. Stoll attributes this strong depth dependence to rapid changes in effective stress (i.e., overburden pressure) with depth. He shows (Stoll, 1989; Fig. 6.6) that attenuation values calculated as a function of mean effective stress, for a thick layer of fine to medium sand, closely agree with attenuations derived from Scholte-waves for the same sand. Indeed, the experimental data presented in Fig. 5 match Stoll’s calculated attenuations (dashed-line profile) in trend and, to some extent, in value (the overall lower magnitude of the dashed profile is consistent with its representation of a sand).

Thus, one can conclude that given accurate sediment information (sediment type being the barest necessity), predicting attenuation from effective stress is probably an acceptable method for general modeling purposes.

Biot theory predicts higher attenuations at higher frequencies (Stoll, 1989) and it can be seen (Fig. 5) that relatively higher frequency (20–40 Hz) borehole measurements follow this relationship. In total, however, the data in Figure 5 do not support this aspect of Biot theory, indicating that more than frequency is involved. Some of the highest attenu-

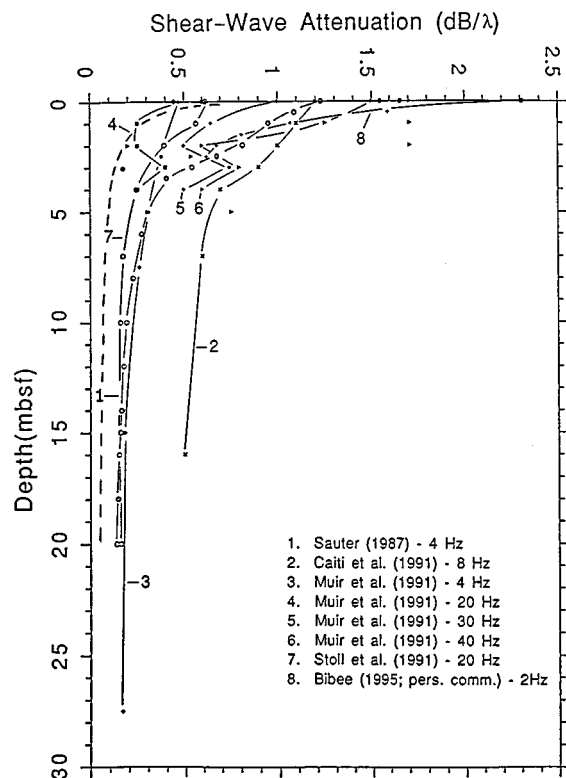


FIG. 5. Expanded scale showing the distribution of shear-wave attenuation measurements for the upper 30 meters of Fig. 4. Lines connect data points pertaining to specific field experiment (as indexed according to the figure legend). Unconnected points represent discrete measurements (solid circles represent sand/silt). Dashed line represents attenuation/depth profile calculated as a function of mean effective stress (from Stoll, 1989).

ation values, belonging to profile 8 (Bibee, personal communication), are based on measurements made at only 2 Hz but in high porosity (90%), low shear velocity (11 m/s), shallow water (250 m), prodelta, silty clay sediments of the Mississippi fan (Bibee and Dorman, 1995). Sediments of this nature are rapidly deposited and tend to be soupy, underconsolidated, low in shear strength (hence, low effective stress and high attenuation). They may also be gassy [although not in this case according to Bibee (personal communication)]. Thus, a discord between theory and observation is predictable when one considers site-to-site variability in sediment physical parameters (e.g., density, porosity) and geologic stress history (e.g., overconsolidation, underconsolidation).

It is evident above that neither theory nor field experiments support Hamilton's (1987) assumption that *S*-wave attenuation is proportional to *P*-wave attenuation. It is also noteworthy that the 20–100 Hz attenuation data presented by Muir *et al.* (1991) varies according to the equation α (dB/m) = $3 \times 10^{-3} f^{1.5}$ and that experimental results reported by Stoll (1991) also indicate a nonlinear relationship. Although these findings conflict with another of Hamilton's (1987) assumptions, i.e., that the relationship between *S*-wave attenuation and frequency is linear, considerably more experimental work is required in order to confirm either point of view.

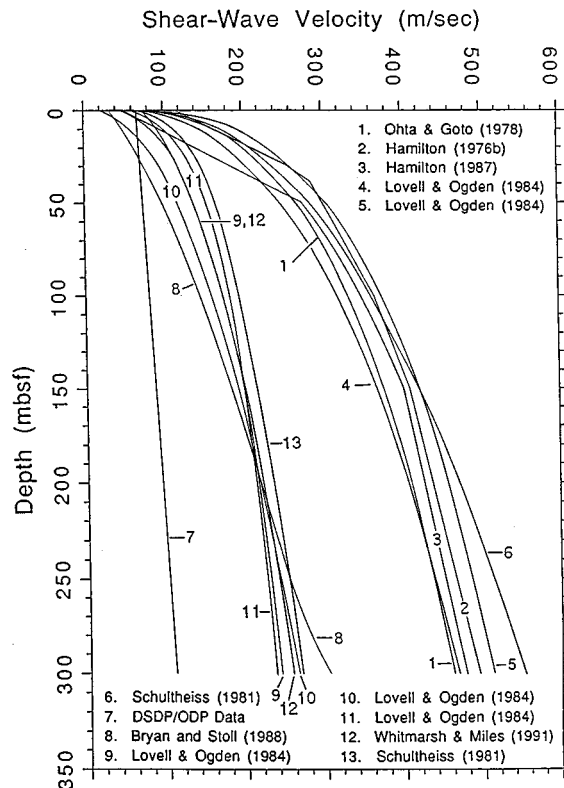


FIG. 6. Shear-wave velocity versus depth profiles derived from published field and laboratory measurements (as indexed according to the figure legend). Profile numbers correspond to the numbered equations in Table III.

C. Shear-wave velocity

Regression equations, empirically derived by Hamilton (1976b, 1987), are commonly used to generate *S*-wave velocity profiles in the upper few hundred meters of sediment. There are, however, a number of published profiles and equations that serve the same purpose but are not widely used, if at all. One reason for this may be that they have never been grouped for comparison and evaluation.

Thirteen profiles are shown in Fig. 6 and their equations listed in Table III. Most of the profiles stem directly or indirectly from measurements made on natural sediments (Table IV). Profile 8, however, is derived from relationships developed by Bryan and Stoll (1988) based on theoretical considerations and requires void ratio as an input. Such models approximate the real world only when the parameter inputs (i.e., values) are themselves realistic. On the other hand, measurement errors and data processing errors also result in unrealistic profiles.

Profiles 6 and 13 are derived from laboratory measurements of *S*-wave velocity made during consolidation testing (Schultheiss, 1981). The resulting regression equations require the input of effective stress, hence, the above cautionary note applies. In order to minimize error, Schultheiss' original velocity/stress profiles were converted to velocity/depth profiles (see Appendix) rather than generate new profiles based on contrived values for effective stress.

With the exception of profile 7, it is evident that the profiles in Fig. 6 fall into two distinct groups (high velocity and low velocity), depending on their initial curvature (i.e.,

TABLE III. Equations for generating *S*-wave velocity/depth profiles. The equations are indexed according to the number given to each curve in Fig. 6 (in all equations *d*=depth in meters).

| | |
|-----|--|
| 1. | $V_s = 78.89d^{0.312}$ |
| 2. | $V_s = 116 + 4.65d$ (where $d = 0-36$ m). $V_s = 237 + 1.28d$ (where $d = 36-120$ m). $V_s = 322 + 0.58d$ (where $d = 120-650$ m). |
| 3. | $V_s = 3.884V_p - 5.757$ (for V_p less than 1.555) $V_s = 1.137V_p - 1.485$ (for V_p from 1.555 to 1.650). $V_s = 0.991 - 1.136V_p + 0.47 V_p^2$ (for V_p from 1.650 to 2.150) V_p is generated from the regression equation given in Hamilton (1987, Table 7). However, the surface sediment velocity was changed from 1.519 to 1.493 km/s which is given by Hamilton (1987, Table 4) as a representative velocity for a pelagic clay. Both V_p and V_s are in km/s. |
| 4. | $V_s = 62.52d^{0.3537}$ |
| 5. | $V_s = 96.80d^{0.29326}$ |
| 6. | $V_s = 0.5\sigma^{0.48} + 80$, where σ =effective stress |
| 7. | $V_s = 77.46 + 0.24079d$ |
| 8. | $V_s = 33.48 + 1.44d + (-3.79 \times 10^{-3})d^2 + 6.67 \times 10^{-6}d^3$ This equation, which will reproduce the profile shown in Fig. 6, is based on the relationships of Bryan and Stoll (1988). |
| 9. | $V_s = 42.71d^{0.3069}$ |
| 10. | $V_s = 20.50d^{0.4496}$ |
| 11. | $V_s = 58.11d^{0.2484}$ |
| 12. | $V_s = 75.4 + 1.51d + (-5.47 \times 10^{-3})d^2 + 8.35 \times 10^{-6}d^3$ |
| 13. | $V_s = 1.2\sigma^{0.37} + 60$, where σ =effective stress |

gradient) near the seafloor. Profile 7 originates from a linear curve-fit of uncorrected, laboratory-measured values extracted from DSDP/ODP reports (see Fig. 8 in Bowles, 1994) and is comparable in magnitude with uncorrected laboratory measurements (maximum sample depth=270 m) made by Schultheiss (1985) and Lavoie and Anderson (1991, Fig. 3). However, neither data set, nor profile 7, exhibit significant velocity gradients. Based on a similar observation, Urmos and Wilkens (1993) concluded that *P*-wave velocity measurements made in the laboratory are of little use unless corrected to *in situ* values. Their conclusion appears to hold for *S*-wave measurements as well.

Profiles 2 and 3 (Fig. 6) are notable, as well as questionable, because they occur as straight lines that cut through the top (0–50 m), convex portions of nearly all the other profiles. Profile 2 tracks with the high-velocity profiles but has an initial velocity (116 m/s) that is unrealistic (i.e., too fast) for most fine-grained, surficial sediment. Profile 3, on

TABLE IV. Summary of methods used to derive shear-wave velocity/depth profiles shown in Figs. 6 and 7.

| Interface waves | Body waves | Empirical or theoretical relationships | Laboratory measurements |
|-----------------|------------|--|-------------------------|
| 14 | 12 | 1 | 4 |
| 15 | | 2 | 5 |
| 16 | | 3 | 6 |
| 17 | | 8 | 7 |
| 18 | | | 9 |
| 19 | | | 10 |
| 20 | | | 11 |
| 21 | | | 13 |
| 22 | | | |
| 23 | | | |

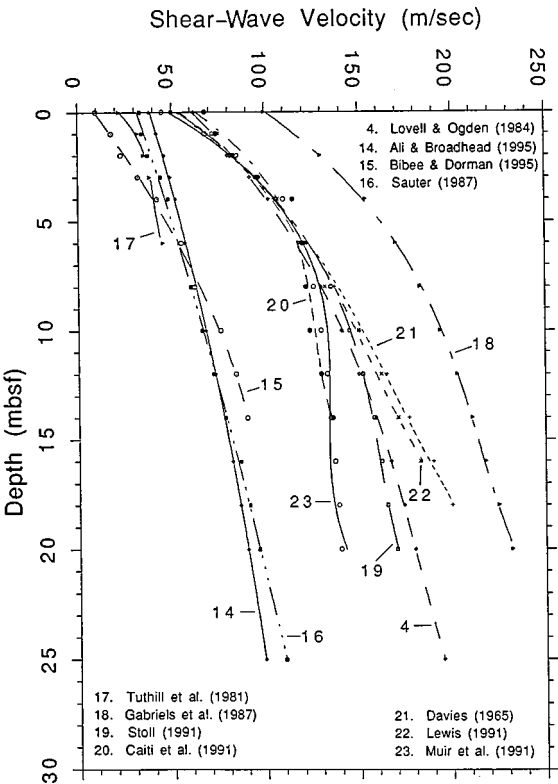


FIG. 7. Shear-wave velocity versus depth profiles derived from published *in situ* measurements made in predominantly fine-grained marine sediments (as indexed according to the figure legend).

the other hand, begins with the low-velocity profiles, leaves the group after about 20 m, and tracks with the high velocity profiles below 50 m.

Most of the profiles (Fig. 6) exhibit substantial increases in *S*-wave velocity in the first 10–20 m of sediment, followed by more gradual increases. These gradients are consistent with velocity gradients found in the upper 20–25 m of muddy sediments by *in situ* measurements (Fig. 7). The *in situ* profiles in Fig. 7 bracket nearly the same range of seafloor velocities (10–116 m/s) as those in Fig. 6 (23–116 m/s), although the average surface velocity in Fig. 7 is lower. Moreover, the *in situ* profiles also fall into two velocity groups that essentially coincide with the velocity groups in Fig. 6 (profile 4 from Fig. 6 has been added to Fig. 7 to show the similarity).

The separation of the profiles (Figs. 6 and 7) into groups is puzzling and one is tempted to seek a reason. An obvious one is measurement technique. This can be ruled out, however, because both velocity groups contain profiles that were obtained from the same technique (Table IV).

Another possible reason is sediment type. Unfortunately, sediment information supporting the profiles is limited and one can only generalize. With regard to Lovell and Ogden's (1984) profiles, 4 and 5 (high-velocity group) relate to carbonate ooze sediments; whereas, 10 and 11 (low-velocity group) relate to red clay sediments. However, profiles 2 and 3 (high-velocity group) are derived from sediments with predominantly terrigenous components (Hamilton, 1976b).

TABLE V. High and low shear-wave velocity profiles (n.d.=no data).

| Profile number | Sediment type | Grain size |
|-----------------------------------|--|---------------------------------------|
| High shear-wave velocity profiles | | |
| 1 | silt/clay soils | n.d. |
| 2 | hemipelagic silts, clays, turbidites | n.d. |
| 3 | hemipelagic silts, clays, turbidites | n.d. |
| 4 | biogenic coarse clay | 8.8 ϕ |
| 5 | biogenic fine silt | 7.2 ϕ |
| 6 | sand-silt-clay | 5.9 ϕ |
| Low shear-wave velocity profiles | | |
| 7 | calcareous clay | n.d. |
| 8 | clayey silt | n.d. |
| 9 | pelagic clays/oozes, fine-grained turbidites | 8.9–10.3 ϕ (ave. 9.4 ϕ) |
| 10 | red clay | 10.9 ϕ |
| 11 | red clay | 10.1 ϕ |
| 12 | pelagic mud/clay | n.d. |
| 13 | potter's clay | 9.4 ϕ |

Hence, differentiation of the profiles based on a terrigenous-biogenic distinction does not work.

There is evidence, however, that the grouping of the profiles is based largely on gross textural differences. Table V summarizes this information and shows (Fig. 6) that the high-velocity profiles represent coarser sediments than the low-velocity profiles. For those profiles lacking quantitative information, (i.e., grain size) sediment types were assigned based upon information gleaned from the literature (see Appendix). The assignments are consistent with the grain size information for each group.

Except for profile 5 (fine silt), Lovell and Ogden's four other profiles (4, 9, 10, and 11) all relate to clay-size sediment. The presence of profile 4 with the high-velocity profiles seems odd in this respect, even though it is the coarsest of the four. This suggests that another factor(s) contributes to the grouping of the profiles. Profiles 4, 5, 10, and 11 are based on samples taken at various sub-bottom depths (25, 79, 45, and 13 m, respectively). Profiles 10 and 11 are both low-velocity profiles, yet they represent deep and shallow sample depths, respectively. Thus, compaction is apparently not a factor. Cementation, on the other hand, is a possibility. At these sample depths, the carbonates (4 and 5), vis-à-vis the noncarbonates (10 and 11), have probably experienced some degree of cementation that has resulted in higher velocities.

Profile 9 is notable because it is based on nine samples with an average median grain-size range of 9.4 ϕ . Logically, one could expect it to occur midway between profiles 4 (8.8 ϕ) and 11 (10.1 ϕ). That it does not may be due to the fact that: (1) it is derived from measurements made on both terrigenous clay and carbonate ooze sediment samples; and, that (2) sample depths (20–300 cm) are too shallow for cementation effects to have developed in the soft oozes. Thus, the sediments for profile 9 behave more like the clays of profiles 10 and 11 than the oozes of profiles 4 and 5.

Textural differences also appear to account for the grouping of the *in situ* profiles in Fig. 7. This is readily

TABLE VI. High and low shear-wave velocity profiles (n.d.=no data).

| Profile number | Sediment type | Grain size |
|-----------------------------------|---|-----------------|
| High shear-wave velocity profiles | | |
| 4 (see Table V) | | |
| 18 | coarse to fine sand, intercalated peat/clay | n.d. |
| 19 | clayey silt | n.d. |
| 20 | sand/silty clay | 5.3–9.7 ϕ |
| 21 | biogenic turbidite sands and ooze | n.d. |
| 22 | hemipelagic silty clay | n.d. |
| 23 | sand/silty clay | 4.0–10.0 ϕ |
| Low shear-wave velocity profiles | | |
| 14 | hemipelagic clay/ooze | n.d. |
| 15 | prodelta clay | n.d. |
| 16 | hemipelagic clay/ooze | n.d. |
| 17 | soft, porous silt | n.d. |

shown by Table VI where, once again, the high-velocity group of profiles correspond with the coarser sediments. As in Table V, the assigned sediment types are based on descriptive, often extrapolated, information (see Appendix).

III. SUMMARY AND CONCLUSIONS

The reliability of geoacoustic models, with respect to the real world, is uncertain because choosing values for the required model inputs is largely an empirical process. Confidence in selecting appropriate values, particularly for *P*-wave attenuation, *S*-wave attenuation, and *S*-wave velocity, is low for two important reasons: (1) existing measurements have been made under wide-ranging conditions (environment, technique, etc.) and, (2) the majority of these measurements are poorly supported by quantitative sediment information (grain size, porosity, etc.). Consequently, it is difficult to interpret experimental results, or compare results from different ocean areas. In this paper, published measurements for the above parameters are compiled for fine-grained sediments and presented as tables and graphs. The latter reveal trends that do not always agree with accepted methods for estimating these parameters.

Presently, the geoacoustic community is largely served by empirical relationships developed by Hamilton (1972, 1976a, b, c, 1980, 1987) who has long advocated a linear (f^1) relationship between *P*-wave attenuation and frequency. A curve fit of *P*-wave attenuation measurements, plotted over 10 Hz to 500 kHz, is best fit with a frequency dependence of $f^{1.12}$. The curve fit conforms closely with theoretical (Biot–Stoll) predictions of attenuation for fine-grained sediment over the same frequency range. It is impossible, however, to statistically differentiate among f^1 , $f^{1.12}$, or the Biot–Stoll model given the variability of the data. Thus, it is appropriate to assume a linear relationship between attenuation and frequency for approximating attenuations from 10 Hz to 500 kHz when modeling the geoacoustic properties of fine-grained (low permeability), unconsolidated sediments.

Hamilton's profile predicting *P*-wave attenuation versus depth is influenced by sediment types that have a significant coarse-grained component. Consequently, his attenuation

values are roughly an order of magnitude higher than usually observed in fine-grained sediments. For reasons discussed, modelers are probably better served by the attenuation-depth profiles of Mitchell and Focke (1980).

The assumption (Hamilton, 1987) that *S*-wave attenuation varies proportionally with *P*-wave attenuation leads to *S*-wave attenuations with unrealistic values. The point is moot, however, because *S*-wave attenuations derived in this manner do not reflect the pattern that is documented by recent data. Attenuations are observed to decrease rapidly in the upper 5 m of sediment; becoming nearly constant at depths below 10–20 m. For purposes of modeling fine-grained sediments at 100 Hz or less, *S*-wave attenuation at the seafloor can be derived using the nonlinear relationship reported by Muir *et al.* (1991). The trends and values observed with depth in recent data can be approximated for modeling purposes from estimates of overburden pressure (i.e., effective stress) as a function of depth.

Other than Hamilton's regression equations for generating *S*-wave velocities in the upper few hundred meters of sediment, the geoaoustic-modeling community seems to have largely ignored the existence of other velocity-depth profiles relating to fine-grained sediments. Composite plots of these profiles (including Hamilton's) reveal an interesting characteristic; they tend to separate into two groups (high velocity and low velocity). Neither measurement technique nor sediment type (i.e., carbonate content) seem to account for the separation; texture does, however. Both qualitative and quantitative evidence indicates that the higher-velocity profiles reflect fine-grained sediments with a coarse-grained component that is not present in the low-velocity profiles.

Hamilton's equations generate linear *S*-wave gradients that appear to be unrepresentative of the upper 50 m of sediment. For this reason alone, many of the other equations (profiles) are probably more suitable for predicting *S*-wave velocities with depth. Choosing the appropriate equation, however, depends on the physical properties of the sediment. Consequently, the best equation may be one that requires the input of a specific sediment parameter (e.g., void ratio or effective stress) as the variable rather than depth beneath the seafloor.

Finally, a better understanding of the variation in *S*-wave velocity with depth (or any acoustic measurement for that matter) would have emerged from this paper had investigators done more to characterize the sediments at their experimental sites. Indeed, the ability of investigators to comprehend and compare field measurements is greatly improved when sediment type is quantified according to a standard nomenclature (e.g., Shepard, 1954) based on sand-silt-clay ratios. Given a sediment type in the same environment in any ocean, excellent empirical relationships exist between sound velocity, grain size, porosity, and wet bulk density (Hamilton and Bachman, 1982; Orsi and Dunn, 1990). At its simplest, obtaining such information could involve no more than searching the literature for areas where bottom samples have already been collected (and suitably analyzed); then, conducting acoustic experiments in these places. In unsampled areas, however, it behooves investigators to collect a suite of sediment cores along geophysical transects for the

combined purposes of determining the above properties as well as the continuity of the sediments within the experimental area. Sampling density, in these cases, will depend on the spatial scale (meters or kilometers) appropriate for the experiment.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Joseph Gettrust, Dr. Dale Bibee, Dr. Peter Fleischer, Dr. Dawn Lavoie, Dr. Dennis Lindwall, Dr. Mike Richardson, and Dr. Warren Wood for critically reviewing the manuscript. This work was supported by the Office of Naval Research-sponsored Acoustic Reverberation Special Research Program, Program Element 0601153N, Project No. R0320404, Program Manager, Dr. Mohsen Badiey; and by the Office of Naval Research through the Naval Research Laboratory-sponsored Bottom Interaction Project, Program Element 0602435N, Project No. BE-35-2-02.

APPENDIX

1. Figure 6

Profile 1: This profile is based on alluvium (Holocene) and diluvium (Pleistocene) soils of Japan. Thus, its association with profiles that relate to marine sediments may seem odd. It is included chiefly because Hamilton (1980, 1987) presents it in his discussions of shear-wave velocity. In their paper, Ohta and Goto (1978) lump both silty and clayey soils into one category (clay), but provide no quantitative information (e.g., grain size, number of silt versus clay soils, etc.). It can be assumed that profile 1, by its association with other high-velocity profiles, is mainly influenced by the silty soils.

Profiles 2 and 3: Both profiles result from empirically derived regression equations based on *S*-wave velocity measurements made on a variety of sediment types. As described (see Table 2 in Hamilton, 1976b), most of these sediments are decidedly terrigenous and often appear to be substantially, if not entirely, coarse grained (sands and silts).

Profiles 3 and 6: These profiles are presented in Schultheiss (1981) but are scaled according to effective stress rather than meters (i.e., sediment depth). In order to convert his stress scale to meters, the following steps were taken: (1) the grain size distributions for his silty clay and potter's clay were matched with average sediment grain size analyses from Hamilton's (1987) tables. Schultheiss' silty clay is, in fact, quite sandy (30%) and compares closely with Hamilton's sand-silt-clay category, while the potter's clay compared well with Hamilton's deep-sea ("red") pelagic clay. Based on these matches, a bulk density was assigned to each of the Schultheiss sediments (silty clay = 1.575 g/cm³; potter's clay = 1.414 g/cm³). For a given sediment depth the stress (σ) was calculated by the equation: $\sigma = \text{depth} \times (\text{sediment bulk density} - \text{water density})$. For each profile the sediment depth was varied until the calculated stress equaled 400 and 800 kPa. Having established these stress/depth reference points, Schultheiss' profiles could then be scaled in meters and transferred to Fig. 6.

Profile 8: This profile is included in Fig. 6 principally because the relationships developed by Bryan and Stoll

(1988) are commonly used for deriving shear velocity/depth profiles. As noted in the text, the relationships require void ratio as an input. In this case, the void ratios are estimated from porosities taken directly from the geoaoustic model developed by Bowles (1994). The porosities range from 68% at the sea floor to 55% at 300 meters and are consistent with the porosity/depth profile of Nobes *et al.* (1986) which is also derived from DSDP porosity data.

Profile 12: Experiments were conducted at 64°15'N/02°17'W but no supporting sediment information is furnished by the investigators for their profile. Fortunately, DSDP site 337 is located just to the north (64°52'N/05°21'W). Here, 47 m of "glacial" mud (roughly 80% clay-size material), with intervals of calcareous ooze, overly 72 m of dominantly (90%) clay (see lithologic descriptions in Talwani *et al.* 1976). Because the sites are close to each other, and both occur on the Norwegian continental rise in similar water depths, there is no compelling reason to believe that the sediments at either site would differ significantly.

2. Figure 7

Profile 15: This profile represents high porosity (90%), muddy, prodelta deposits of the Mississippi fan (Bibee and Dorman, 1995). Such deposits characteristically consist of fine-grained muddy sediments, i.e., clay and silty clay; away from the delta front, clays predominate (Reineck and Singh, 1975). Extremely low *S*-wave velocities (7–9 m/s) have also been measured in high porosity (91%–84%), fine-grained (9.9 ϕ) sediments by Richardson and Briggs (1996).

Profiles 14 and 16: Both profiles reflect the velocity structure of the sediments near DSDP Site 469 where the upper 42 m of sediment is extremely fine grained, consisting of hemipelagic clay with intermixed foram-nannofossil ooze (see lithologic descriptions in Yeats *et al.*, 1981). Most encouraging is the observation that the profiles nearly overlap one another, as they should.

Profile 17: The profile is anomalous if one assumes that coarseness generally dictates higher sediment velocity. The profile represents an acoustically transparent (i.e., structureless) sediment that is 94% silt (Sutton *et al.*, 1981); yet, it falls with the low-velocity profiles. Trehu and Solomon (1981) describe the deposit as very soft, porous, and organic, such that a core sampler could be pushed in by hand. Thus, the low velocities apparently result from low sediment rigidity (shear strength) annulling the effect of grain size. Clearly, it is desirable to have more than one sediment parameter (i.e., grain size) at hand when making interpretations.

Profile 18: This profile is the most noticeable because it stands alone. The high velocities it portrays result from the fact that the seismic experiment was conducted in a tidal flat deposit consisting mostly of coarse to fine sand (Gabriels *et al.*, 1987).

Profile 19: The experimental site for this profile was situated in the Hudson River (New York) just north of the Tappan Zee bridge where the bottom consists of 10–30 m of soft, organic silt overlying a thick layer of silty clay (Stoll, 1991; Stoll *et al.*, 1991).

Profile 21: The only information furnished with this profile is the experimental site (02°11'S, 57°23'E), which is relatively close to DSDP Site 236 (1°41'S, 57°39'E). In the latter case, the sediments are mainly foram-nannofossil oozes (see lithologic descriptions in Fisher *et al.*, 1974) which would suggest a profile with lower shear velocities. The drilling process, however, is notable for not recovering surficial sediments (where Scholte waves travel). Of two standard sediment cores taken nearby (2°27'S, 53°37'E and 2°40'S, 54°45'E), laboratory analysis of one (Naval Oceanographic Office) discloses a combined sand/silt content of 30%–40% for the top 5 m. For the other core, unpublished megascopic descriptions (Lamont–Doherty Geological Observatory) reveal about 16 m of foram-nannofossil ooze interlayered with numerous, thick (up to 135 cm) foram sand layers. The sands were presumably deposited by turbidity currents originating on the slopes of nearby Seychelles Bank and the same depositional process most likely characterizes the experimental area of profile 21 as well. Whitmarsh and Lilwall (1982) used the same interpretational technique as Davies (1965, profile 21) to produce an *S*-wave velocity/depth profile for the Madeira abyssal plain that is remarkably similar to Davies' profile (hence it is not shown in Fig. 7). Their examination of sediment cores taken from the Madeira abyssal plain, and within 6 km of Davies' Indian Ocean site, showed that both areas consist of nanno-foram oozes with interbedded, graded, biogenic, turbidite layers.

Profile 22: The experimental site (44°36'N, 125°16'W) is described as a continental slope-basin (2200 m) off the Oregon coast (Lewis, 1991). However, DSDP Site 175 is located just to the north (44°50'N, 125°14'W) in 2000 m of water. Here the upper 120 m of sediment consist of gray, hemipelagic, silty clay. According to Hamilton (1980, 1987), a typical continental terrace silty clay will have a silt content of 35%–41% and a mean grain size of 8.5–8.8 ϕ .

- Anderson, R. S., and Blackman, A. (1971). "Attenuation of low-frequency sound waves in sediment," *J. Acoust. Soc. Am.* **49**, 786–791.
- Ali, B. H., and Broadhead, M. K. (1995). "Shear wave properties from inversion of Scholte wave data," in *Full Field Inversion Methods in Ocean and Seismic Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht), pp. 371–376.
- Barker, J. S., and Helmberger, D. V. (1986). "A broad-band study of attenuation in ocean bottom sediments," in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York), pp. 609–621.
- Bennett, Jr., L. C. (1966). "In situ measurements of acoustic absorption in unconsolidated marine sediments," Ph.D. dissertation, Bryn Mawr College.
- Berzon, I. S., Batnikova, L. I., and Mitronova, V. A. (1969). "Shear-waves reflected from a thin high-speed layer," in *Shear Wave Propagation in Real Media*, edited by I. S. Berzon (Consultants Bureau, New York), pp. 154–162.
- Bibee (1995). Personal communication.
- Bibee, L. D., and Dorman, L. H. (1991). "Implications of deep-water seismometer array measurements for Scholte wave propagation," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht), pp. 221–230.
- Bibee, L. D., and Dorman, L. M. (1995). "Full waveform inversion of seismic interface wave data," in *Full Field Inversion Methods in Ocean and Seismic Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht), pp. 377–382.
- Bowles, F. A. (1994). "A geoaoustic model for fine-grained, unconsolidated calcareous sediments (ARSRP Natural Laboratory)," Naval Research Laboratory, Stennis Space Center, MS, Memorandum Report No. NRL/MR/7432–93-7082, 55 pp.

- Brienzo, R. K. (1992). "Velocity and attenuation profiles in the Monterey Deep-Sea Fan," *J. Acoust. Soc. Am.* **92**, 2109–2125.
- Bromirski, P. D., Frazer, L. N., and Duennebie, F. K. (1992). "Sediment shear Q_β from airgun OBS data," *Geophys. J. Int.* **110**, 465–485.
- Bryan, G. M., and Stoll, R. D. (1988). "The dynamic shear modulus of marine sediments," *J. Acoust. Soc. Am.* **83**, 2159–2164.
- Bucker, H. P., Whitney, J. A., and Keir, D. L. (1964). "Use of Stoneley waves to determine the shear velocity in ocean sediments," *J. Acoust. Soc. Am.* **36**, 1595–1596.
- Caiti, A., Akal, T., and Stoll, R. D. (1991). "Determination of shear velocity profiles by inversion of surface wave data," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht), pp. 557–565.
- Carter, J. A., Sutton, G. H., and Setau-Henson, A. (1986). "Analysis of ocean-subbottom seismograph (OSS) data," in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York), pp. 553–563.
- Caruthers, J. W., Bowles, F. A., and Kalra, A. K. (1991). "Site characterization for the ARSRP scattering experiment," *J. Acoust. Soc. Am.* **89**, 1964.
- Chapman, N. R. (1980). "Low frequency bottom reflectivity measurements in the Tufts Abyssal Plain," in *Bottom-Interacting Ocean Acoustics*, edited by W. A. Kuperman and F. B. Jensen (Plenum, New York), pp. 193–207.
- Chotiros, N. P. (1989). "High frequency acoustic penetration analysis," Applied Research Laboratories, University of Texas (Austin), Technical Report No. ARL-TR-89-28, 122 pp.
- Courtney, R. C., and Mayer, L. (1993). "Acoustic properties of fine-grained sediments from Emerald Basin: Toward an inversion for physical properties using the Biot-Stoll model," *J. Acoust. Soc. Am.* **93**, 3193–3200.
- Davies, D. (1965). "Dispersed Stoneley waves on the ocean bottom," *Bull. Seis. Soc. Am.* **55**, 903–918.
- Dicus, R. L. (1976). "Preliminary investigations of the ocean bottom impulse response at low frequencies," Naval Oceanographic Office, Stennis Space Center, MS, Technical. Note 6130-4-76.
- Dicus, R. L., and Anderson, R. S. (1982). "Effective low-frequency geoaoustic properties inferred from measurements in the northeast Atlantic," Naval Research Laboratory, Stennis Space Center, MS, NORDA Report No. 21, 155 pp.
- Duckworth, G. L., and Baggeroer, A. B. (1986). "Estimation of ice surface scattering and acoustic attenuation in Arctic sediments from long-range propagation data," in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York), pp. 373–386.
- Ferla, M. C., Dreini, F. B., Jensen, F. B., and Kuperman, W. A. (1980). "Broadband model/data comparisons for acoustic propagation in coastal waters," in *Bottom-Interacting Ocean Acoustics*, edited by W. A. Kuperman and F. B. Jensen (Plenum, New York), pp. 577–592.
- Fisher, R. L., Bunce, E. T., Cernock, P. J., Clegg, D. C., Cronan, D. S., Damiani, V. V., Dmitriev, L. V., Kinsman, D. J. J., Roth, P. H., Thiede, J., and Vincent, E. (1974). *Initial Reports of the Deep Sea Drilling Project, Vol. 24* (U.S. Government Printing Office, Washington, DC), pp. 327–389.
- Focke, K. C. (1983). "Acoustic attenuation in ocean sediments found in shallow water regions," Applied Research Laboratories, University of Texas (Austin) Technical Report No. ARL-TR-84-6.
- Frisk, G. V., Douth, J. A., and Hays, E. E. (1981). "Bottom interaction of low-frequency acoustic signals at small grazing angles in the deep ocean," *J. Acoust. Soc. Am.* **69**, 84–94.
- Frisk, G. V., Douth, J. A., and Hays, E. E. (1986). "Geoaoustic models for the Icelandic Basin," *J. Acoust. Soc. Am.* **80**, 591–600.
- Gabriels, P., Snieder, R., and Nolet, G. (1987). "In situ measurements of shear-wave velocity in sediments with higher-mode Rayleigh waves," *Geophys. Prospect.* **35**, 187–196.
- Hamilton, E. L. (1972). "Compressional-wave attenuation in marine sediments," *Geophysics* **37**, 620–646.
- Hamilton, E. L. (1974). "Prediction of deep-sea sediment properties: state-of-the-art," in *Deep Sea Sediments: Physical and Mechanical Properties*, edited by A. L. Inderbitzen (Plenum, New York), pp. 1–43.
- Hamilton, E. L. (1976a). "Sound attenuation as a function of depth in the deep sea floor," *J. Acoust. Soc. Am.* **59**, 528–535.
- Hamilton, E. L. (1976b). "Shear wave velocity versus depth in marine sediments: a review," *Geophysics* **41**, 985–996.
- Hamilton, E. L. (1976c). "Attenuation of shear waves in marine sediments," *J. Acoust. Soc. Am.* **60**, 334–338.
- Hamilton, E. L. (1980). "Geoacoustic modeling of the sea floor," *J. Acoust. Soc. Am.* **68**, 1313–1340.
- Hamilton, E. L. (1987). "Acoustic properties of sediments," in *Acoustics and Ocean Bottom*, edited by A. Lara-Saenz, C. Ranz-Guerra, and C. Caro-Fite (Sociedad Espanola de Acustica, Instituto de Acustica-CSIC, F. A. S. E. Specialized Conference, Madrid), pp. 3–58.
- Hamilton, E. L., and Bachman, R. T. (1982). "Sound velocity and related properties of marine sediments," *J. Acoust. Soc. Am.* **72**, 1891–1904.
- Hampton, L. (1985). "Acoustic properties of sediments: an update," *Rev. Geophys.* **23**, 49–60.
- Helmberger, D. V., Engen, G., and Scott, P. (1979). "A note on velocity, density, and attenuation models for marine sediments determined from multibounce phases," *J. Geophys. Res.* **84**, 667–671.
- Holt, R. M., Hovem, J. M., and Syrtstad, J. (1983). "Shear modulus profiling of near bottom sediments using boundary waves," in *Acoustics and the Sea-Bed*, edited by N. G. Pace (Bath U.P., Bath, UK), pp. 317–325.
- Hooper, M. S., Ingram, G. D., and Mitchell, S. K. (1981). "Measurements and analysis of acoustic bottom interaction in the northwestern Mexican basin," Technical Report No. ARL-TR-81-37, Applied Research Laboratories, University of Texas (Austin).
- Jacobson, R. S., Shor, Jr., G. G., and Bee, M. (1984). "A comparison of velocity and attenuation between the Nicobar and Bengal Deep Sea Fans," *Geophysics* **89**, 6181–6196.
- Jacobson, R. S., Shor, Jr., G. G., and Dorman, L. M. (1981). "Linear inversion of body wave data-Part II: Attenuation versus depth using spectral ratios," *Geophysics* **46**, 152–162.
- Jensen, F. B., and Schmidt, H. (1986). "Shear properties of ocean sediments determined from numerical modelling of Scholte wave data," in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York), pp. 683–692.
- Joyner, W. B., Warrick, R. E., and Oliver III, A. A. (1976). "Analysis of seismograms from a downhole array in sediments near San Francisco Bay," *Bull. Seismol. Soc. Am.* **66**, 937–958.
- Kibblewhite, A. C. (1989). "Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data," *J. Acoust. Soc. Am.* **86**, 716–738.
- Lavoie, D., and Anderson, A. (1991). "Laboratory measurements of acoustic properties of periplatform carbonate sediments," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht), pp. 111–120.
- Lewis, B. T. R. (1991). "Changes in P and S velocities caused by subduction related sediment accretion off Washington/Oregon," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht), pp. 379–386.
- Liu, H.-P., Warrick, R. E., Westerlund, R. E., and Kayen, R. E. (1994). "In situ measurement of seismic shear-wave absorption in the San Francisco Holocene Bay mud by the pulse-broadening method," *Bull. Seism. Soc. Am.* **84**, 62–75.
- Lovell, M. A., and Ogden, P. (1984). "Remote assessment of permeability/thermal diffusivity of consolidated clay sediments," Final Report (EUR 9206 en) to Commission of the European Communities, Nuclear Science and Technology, Luxembourg, 168 pp.
- Matthews, J. E. (1982). "Geoacoustic models for the Straits of Sicily and Sardinia-Tunisia," Naval Research Laboratory, Stennis Space Center, MS, NORDA Technical Note 98, 74 pp.
- McCann, C., and McCann, D. M. (1969). "The attenuation of compressional waves in marine sediments," *Geophysics* **34**, 882–892.
- McLeroy, E. G., and DeLoach, A. (1968). "Sound speed and attenuation, from 15 to 1500 kHz, measured in natural sea-floor sediments," *J. Acoust. Soc. Am.* **44**, 1148–1150.
- Mitchell, S. K., and Focke, K. C. (1980). "New measurements of compressional wave attenuation in deep ocean sediments," *J. Acoust. Soc. Am.* **67**, 1582–1589.
- Muir, T. G., Akal, T., Richardson, M. D., Stoll, R. D., Caiti, A., and Hovem, J. M. (1991). "Comparison of techniques for shear wave velocity and attenuation measurements," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht), pp. 283–294.
- Neprochnov, Yu. P. (1971). "Seismic studies of the crustal structure beneath the seas and oceans," *Oceanology* **11**, 709–715.
- Nobes, D. C., Villinger, H., Davis, E. E., and Law, L. K. (1986). "Estimation of marine sediment bulk density properties at depth from seafloor geophysical measurements," *J. Geophys. Res.* **91**, 14 033–14 043.
- Ohta, Y., and Goto, N. (1978). "Empirical shear wave velocity equations in

- terms of characteristic soil indexes," *Earthquake Eng. Structural Dynamics* **6**, 167.
- Orsi, T. H., and Dunn, D. A. (1990). "Sound velocity and related physical properties of fine-grained abyssal sediments from the Brazil Basin (South Atlantic Ocean)," *J. Acoust. Soc. Am.* **88**, 1536–1542.
- Rauch, D. (1986). "On the role of bottom interface waves in ocean seismoacoustics: a review," in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York), pp. 623–641.
- Reineck, H.-E., and Singh, I. B. (1975). *Depositional Sedimentary Environments* (Springer-Verlag, New York), 439 pp.
- Richardson, M. D., and Briggs, K. B. (1996). "In situ and laboratory geoaoustic measurements in soft mud and hard-packed sand sediments: Implications for high-frequency acoustic propagation and scattering," *Geo. Mar. Lett.* **16**, 196–203.
- Richardson, M. D., Curzi, P. V., Muzi, E., Turgutcan, F., and Akal T. (1992). "A generic geoaoustic model for the central Adriatic Sea," SACLAN Undersea Research Centre, La Spezia, Italy, Report No. SR-199, 89 pp.
- Rogers, A. K., Yamamoto, T., and Carey, W. (1993). "Experimental investigation of sediment effect on acoustic wave propagation in the shallow ocean," *J. Acoust. Soc. Am.* **93**, 1747–1761.
- Rubano, L. A. (1980). "Acoustic propagation in shallow water over a low-velocity bottom," *J. Acoust. Soc. Am.* **67**, 1608–1613.
- Sauter, A. W. (1987). "Studies of the upper oceanic floor using ocean bottom seismometers," Ph.D. dissertation, University of California (San Diego), 100 pp.
- Schirmer, F. (1971). "Eine untersuchung akustischer eigenschaften von sedimenten der Nord- und Ostsee," Ph.D. dissertation, University of Hamburg.
- Schirmer, F. (1980). "Experimental determination of properties of the Scholte wave in the bottom of the North Sea," in *Bottom Interacting Ocean Acoustics*, edited by W. A. Kuperman and F. B. Jensen (Plenum, New York), pp. 285–298.
- Schultheiss, P. J. (1981). "Simultaneous measurement of *P* and *S* wave velocities during conventional laboratory soil testing procedures," *Mar. Geotechnol.* **4**, 343–367.
- Schultheiss, P. J. (1985). "Physical and geotechnical properties of sediments from the northwest Pacific," in *Initial Reports of the Deep-Sea Drilling Project*, 86, edited by G. R. Heath *et al.* (U.S. Government Printing Office, Washington, DC), pp. 701–722.
- Sereno, T. J. Jr. (1990). "Numerical modeling of water waves recorded by a subbottom seismometer in the southwest Pacific," *J. Geophys. Res.* **95**, 2575–2591.
- Shepard, F. P. (1954). "Nomenclature based on sand-silt-clay ratios," *J. Sediment. Petrol.* **24**, 151–158.
- Shumway, G. (1960). "Sound speed and absorption studies of marine sediments by a resonance method. Part I," *Geophysics* **25**, 451–467.
- Stoll, R. D. (1985). "Marine sediment acoustics," *J. Acoust. Soc. Am.* **77**, 1789–1799.
- Stoll, R. D. (1989). *Sediment Acoustics* (Springer-Verlag, New York), 153 pp.
- Stoll, R. D. (1991). "Shear waves in marine sediments—bridging the gap from theory to field applications," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht), pp. 3–12.
- Stoll, R. D., and Bryan, G. M. (1970). "Wave attenuation in saturated sediments," *J. Acoust. Soc. Am.* **47**, 1440–1447.
- Stoll, R. D., and Houtz, R. E. (1983). "Attenuation measurements with sonobuoys," *J. Acoust. Soc. Am.* **73**, 163–172.
- Stoll, R. D., Bryan, G. M., Mithal, R., and Flood, R. (1991). "Field experiments to study seafloor seismoacoustic response," *J. Acoust. Soc. Am.* **89**, 2232–2240.
- Sutton, G. H., Lewis, B. T. R., Ewing, J., Duennebie, F. K., Iwatake, B., and Tuthill, J. D. (1981). "An overview and general results of the Lopez Island OBS experiment," *Mar. Geophys. Res.* **5**, 3–34.
- Talwani, M., Udintsev, G., Bjorklund, K., Caston, V. N. D., Faas, R. W., Kharin, G. N., Morris, D. A., Muller, C., Nilsen, T. H., van Hinte, J., Warnke, D. A., and White, S. M. (1976). *Initial Reports of the Deep Sea Drilling Project*, 38 (U.S. Government Printing Office, Washington, DC), 1256 pp.
- Tang, D., Jin, G., Jackson, D. R., and Williams, K. L. (1994). "Analyses of high-frequency bottom and subbottom backscattering for two distinct shallow water environments," *J. Acoust. Soc. Am.* **96**, 2930–2936.
- Tindle, C. T. (1982). "Attenuation parameters from normal mode measurements," *J. Acoust. Soc. Am.* **71**, 1145–1148.
- Trehu, A. M., and Solomon, S. C. (1981). "Coupling parameters of the MIT OBS at two nearshore sites," *Mar. Geophys. Res.* **5**, 69–78.
- Tucholke, B. E., Macdonald, K. D., and Fox, P. J. (1991). "ONR seafloor Natural Laboratories on slow- and fast-spreading mid-ocean ridges," *EOS Trans. Am. Geophys. Union* **72**, 268–270.
- Tuthill, J. D., Lewis, B. T. R., and Garmany, J. D. (1981). "Stoneley waves, Lopez Island noise, and deep sea noise from 1 to 5 Hz," *Mar. Geophys. Res.* **5**, 95–108.
- Tyce, R. C. (1976). "Near-bottom observations of 4 kHz acoustic reflectivity and attenuation," *Geophysics* **41**, 673–699.
- Tyce, R. C. (1981). "Estimating acoustic attenuation from a quantitative seismic profiler," *Geophysics* **46**, 1364–1378.
- Ulonka, A. (1968). "Versuche zur messung der schallgeschwindigkeit und schalldämpfung im sediment in situ," *Deut. Hydrogr. Z., Jahr. 2, Heft 2*, 49–58.
- Urmos, J., and Wilkens, R. H. (1993). "In situ velocities in pelagic carbonates: new insights from Ocean Drilling Program leg 130, Ontong Java plateau," *J. Geophys. Res.* **98**, 7903–7920.
- Warrick, R. (1974). "Seismic investigation of a San Francisco Bay mud site," *Bull. Seismol. Soc. Am.* **64**, 375–385.
- Whitmarsh, R. B., and Lilwall, R. C. (1982). "A new method for the determination of in-situ shear-wave velocity in deep-sea sediments," *Oceanology Int., Paper No. 4.2*, 21 pp.
- Whitmarsh, R. B., and Miles, P. R. (1991). "In situ measurements of shear-wave velocity in ocean sediments," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht), pp. 321–328.
- Wood, A. B., and Weston, D. E. (1964). "The propagation of sound in mud," *Acustica* **14**, 156–162.
- Wroldstad, K. (1980). "Interval velocity and attenuation measurements in sediments from marine seismic reflection data," *J. Acoust. Soc. Am.* **68**, 1415–1435.
- Yeats, R. S., Haq, B. U., Barron, J. A., Bukry, D., Crouch, J., Denham, C., Douglas, A. G., Grechin, V. I., Leinen, M., Niemi, A., Verma, S. P., Pisciotto, K. A., Poore, R. Z., Shibata, T., and Wolfart, R. (1981). *Initial Reports of the Deep Sea Drilling Project*, 63 (U.S. Government Printing Office, Washington, DC), pp. 173–226.

Direct laboratory measurement of forward scattering by individual fish

Li Ding

National Research Institute of Fisheries Engineering, Hasaki, Kashima, Ibaraki, 314-04 Japan

(Received 18 May 1996; revised 30 July 1996; accepted 2 November 1996)

Acoustic backscatter has been widely used to detect and quantify fish. However, other types of information can also be exploited for this purpose. One is acoustic signals scattered by fish in the forward direction. This paper describes a novel laboratory experiment aimed at direct measurement of forward scattering by individual fish. A copper sphere is first used to test the measurement concept and establish experimental procedures. Then forward scattering by fish, in both the dorsal aspect and side aspect, is measured against the incident angle of sound, and simultaneous measurements of backscatter are also made. The results indicate that the forward scattering of individual fish is significantly stronger than the backscattering: At or near normal incidence, the forward scattering is 14–17 dB stronger. It is also found that the forward scattering varies with the incident angle of sound to a much less extent than the backscattering. These features could prove to be the advantages of using forward scatter for detection and quantification of fish. How to make the best use of forward-scattered signals remains to be exploited. © 1997 Acoustical Society of America. [S0001-4966(97)00405-0]

PACS numbers: 43.30.Gv, 43.30.Sf, 43.30.Xm [JHM]

INTRODUCTION

Underwater sound has long been used to detect fish at sea, and more recently, to estimate fish abundance, thus playing an important role in survey and management of global fish stocks (MacLennan and Simmonds, 1992). The information on which acoustic techniques are based has been the sound backscattered from fish. There has been great progress in this area, but there are still some obvious limitations in various conditions (ICES, 1996). While echo-based techniques continue to be an important tool in fisheries sciences, one can also exploit different types of acoustic information for detection of fish. One possible candidate is sound scattered by fish in the forward direction, which in principle can be detected with a transmitter on one side of the target and a receiver on the other side. As a matter of fact, in human-kind's earliest attempt to use underwater sound for fish detection, it was forward scatter, rather than backscatter, that was used. Kimura (1929) placed a transmitter and a receiver in a 43×28 m pond with a depth of 1.5–4.5 m, near the two end sides (spacing 43 m). The beam directions were adjusted such that the transmitted waves, after reflection at one of the side walls, were well received (see also MacLennan and Simmonds, 1992). When fish in the pond passed across the acoustic path, Kimura found that the received wave amplitude noticeably fluctuated. While this was only a qualitative study, forward scatter was implicitly used.

However, forward scatter was quickly replaced by backscatter in later development of fisheries acoustics. It is not until recently that forward scatter has been given significant attention. In water current speed data collected with an acoustic scintillation flowmeter deployed in the south arm of the Fraser River in British Columbia, Canada (Curran *et al.*, 1994), anomalous but reasonably consistent upstream speeds were observed even when the water flow was downstream. Curran *et al.* speculate that the anomalous upstream speeds

were due to the forward scattering by migrating salmon. A question thus arises whether forward scatter can be used for the purpose of detecting and quantifying fish.

There are a few possible techniques that can make use of forward scatter. One is acoustic scintillation which relies on spatial and temporal variability of sound forward scattered by inhomogeneities in water such as variation in the refractive index, and which has been used to measure current speed and turbulence (Farmer and Clifford, 1986). The passage of fish would modify the variability of the signal, from which information about the fish can be inferred (Curran *et al.*, 1994; Ye, 1995). Another possible approach is to use reciprocal transmission, which involves the use of a transmitter and a receiver facing each other, and analysis of times of arrival of transmitted signals. If there are fish moving across the acoustic path, the time of arrival would be altered from which the fish could be detected (Ye and Farmer, 1996b). A third method is by measuring attenuation of sound penetrating through an aggregation of fish. Measurements of the signal attenuation have been obtained to infer the extinction cross section (proportional to attenuation) of fish (Furusawa *et al.*, 1992), which is related to the imaginary part of the forward-scattering function based on the forward-scattering theorem (Ishimaru, 1978). Ye (1996) has used the forward-scattering theorem and a deformed cylinder scattering model (Stanton, 1989) to calculate extinction cross section of fish and found that the results compared favorably with experiments. If the forward-scattering function is appropriately modeled, fish density can be derived from attenuation measurements.

Use of forward scatter may enjoy a few advantages. First, as will be seen below, the forward scattering by fish is much stronger than the backscattering. Second, as pointed out by Ye and Farmer (1996a), the forward scattering of fish may be dominated by the contribution from the fish body at

higher frequencies, unlike the backscattering in which the swimbladder is the dominant factor. This benefits theoretical modeling as well as measurement, since the shape of a swimbladder changes from time to time whereas a fish body remains relatively unchanged. Third, the forward scattering of fish varies with the angle of incidence of sound less significantly than backscattering as suggested by our experimental results. In real situations where fish swim and the angle of incidence changes from time to time, the relative insensitivity should increase the stability of acoustic measurements and add confidence to statistical inference of fish information from the acoustic data.

Just as knowledge of the backscattering cross section of fish is important for reliable estimation of fish abundance with echo sounders, knowledge of the forward-scattering characteristics is essential for development and application of techniques based on the forward-scattering principle, and therefore measurement in this respect is imperative. To the author's best knowledge, measurements of forward scatter of single fish have not appeared in literature. Such measurements are not only essential for development of novel fish detection techniques, but also important for a more complete understanding of the physics of sound scattering by fish.

In this paper, we describe a laboratory experiment designed to measure forward scatter of individual fish, and present some preliminary results. These results clearly indicate that the forward scattering is significantly stronger than the backscattering. We also investigate the variability of forward scattering with respect to the incident angle of sound. Considering possible application in shallow water with horizontally directed sonar, the side-aspect scattering has been of particular interest in our study.

I. EXPERIMENT

A. Experimental setup

The experiment was carried out in the water tank of the National Research Institute of Fisheries Engineering (NRIFE) of Japan, in the spring of 1996. The tank is 15 m long, 10 m wide and 10 m deep. One of the 10-m \times 10-m end walls has wedge-shaped sound absorbers made of rubber with air cavities inside. During the period of the experiment, the tank was filled with fresh water and the water depth was 9.25 m.

Figure 1 shows a diagram of the experimental setup. A split-beam transducer is mounted on the floor of the tank, looking upwards to the water surface. It is operated at 38 kHz by a versatile echo sounding system (model KJ1000), as described in Furusawa *et al.* (1993). The transducer has a radius of 18 cm and thus its 3-dB beamwidth is 6.2° at 38 kHz and sound speed of 1500 ms⁻¹ in the narrow beam mode.

The NRIFE has designed a motor-driven rotating system, which consists of a motor and a rotating bar, nylon (or steel) strings, a number of pulleys, and two weights to maintain tension on the strings (Fig. 1). A target fish is suspended between the two vertical strings of the rotating system using a thin nylon string and two hooks on the tail and upper jaw of the fish. A small float is used to help stabilize the fish in a

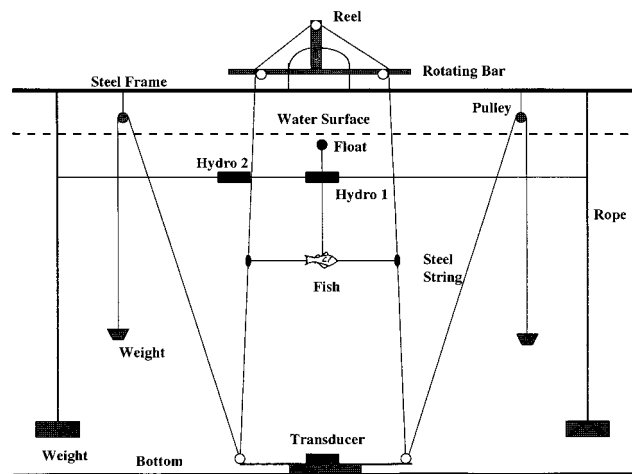


FIG. 1. Experimental setup (not scaled with real dimensions). There are two suspension frames. (1) Suspension frame for hydrophones, including ropes, weights, and nylon string. (2) Suspension frame for fish, including steel strings, pulleys, and weights. The fish is 5.15–5.21 m and the hydrophones are 7.69 m, over the transducer. The water depth is 9.25 m.

desired aspect. Two ropes with heavy weights at the ends are tied to the steel frame which spans the tank, and placed in the water. Two hydrophones are suspended between the ropes at the same depth which is 7.69 m away from the transducer. The first hydrophone is over the fish, and is used to detect forward scatter. Special care is taken to align the fish and this hydrophone on the beam axis of the transducer. This can be done with the help of the split-beam outputs of the transducer. The second hydrophone is about 0.96 m away from the first one, and is positioned at the null of the beam pattern of the transducer. This is to receive minimum direct-path signal amplitude so as to see scattered signals more clearly.

During the experiment, the fish was placed at 5.15–5.21 m above the transducer, and the fork length ranged from 28 to 34 cm. So the entire fish was in the far field and well within the first Fresnel zone (Clay and Medwin, 1977) (also within the 3-dB beam) when the rotating bar was horizontal. When the bar rotates, the fish is tilted relative to the beam axis, and this is equivalent to changing the incident angle of sound. The rotating system is controlled by a computer, and the tilt angle can be incremented by 1°, ranging from –50° to +50° (positive angles correspond to the head towards the source transducer). It is found that when the fish is tilted, its center may move away from the axis. However, this can be monitored and calibrated using the split beam.

B. Experimental procedure and data recording

The most significant difficulty of measuring forward scatter is that the forward-scattered signal arrives at the hydrophone almost at the same time as the direct path signal, in contrast to measuring backscatter in which case the echo can be separated from the transmitted signal in time with a ping-pong technique. In the case of narrow-band transmission, detecting forward-scattered signals is equivalent to detecting a weak sinusoidal signal embedded in a much stronger sinusoidal signal with some phase difference. In real applications this would present a challenging task for signal design and processing. However, in well-controlled laboratory condi-

tions, we can simply record first the waveform received at the hydrophone in the absence of fish, which is the free-field signal. Then we record the received signal with the fish in the water, from which we can subtract the free-field signal. If the position of the hydrophone relative to the transducer is unchanged and the medium (water) remains steady, then the difference can be considered the forward scattered signal. We collected data in such a procedure using a copper sphere as the target and found the result consistent with theory. Such coherent subtraction is, however, very sensitive to any motion of the hydrophone. In the two measurements (with and without a target), the recorded waveforms should not be affected artificially on the phase. To acquire a stable waveform, for each measurement we averaged over a number of pings, and also took special care to keep the hydrophone stable.

Bistatic scattering is measured with the second hydrophone following the same procedure as described. Since the direct path signal arriving at this hydrophone is much weaker, measurements are less sensitive. Simultaneously, the echo signal is also received with the narrow beam of the transducer for comparison with the forward-scattering measurement.

As mentioned above, the split beam is also operated to monitor the position of the fish. We found that at larger tilt angles, the waveform backscattered from the fish deteriorated, possibly due to interference of echoes from different parts of the fish, and that the measured fish position was not consistent with visual observations. This problem itself deserves a further examination, but at present we use echoes from the float to determine the fish position, since the float is tied to the fish center. In the forward-scattering measurement, we place the float very close to the surface to prevent the resulting echo from interfering with the measurement. After each measurement, we redeploy the fish to the same depth and move the float closer to the fish for calibration of the fish position.

Received signals are monitored and recorded with a digital oscilloscope (Lecroy 9304AM). The oscilloscope can simultaneously display waveforms of up to 4 channels, and record them on to hard disks for further analysis. It has capacity of 10 giga-samples/second repetitive sampling, thus providing a very high time resolution. The averaging and subtraction can also be performed internally on the oscilloscope, allowing quick inspection of the data during the experiment.

II. MEASUREMENT PRINCIPLES

Given our present experimental setup, the forward-scattering amplitude of a target can be derived straightforwardly. If the receiving hydrophone and the target are aligned on the axis of the transducer, the beam pattern of the transducer can be omitted (equal to unity), and the incident wave at the target (p_i) and the direct path wave received at the first hydrophone (p_d) can be written as

$$p_i = \frac{P_0}{r_i} e^{-jkr_i}, \quad p_d = \frac{P_0}{r_d} e^{-jkr_d}, \quad (1)$$

and the forward-scattered wave in the far-field (p_f) as

$$p_f = \frac{p_i}{r_1} F_1 e^{-jkr_1} = \frac{P_0}{r_i r_1} F_1 e^{-jk(r_1+r_i)}, \quad (2)$$

where r_i and r_d are the distance from the transducer to the target and to the first hydrophone, respectively, and r_1 is the distance from the target to the first hydrophone, i.e., $r_d = r_i + r_1$. P_0 is the pressure amplitude at 1-m distance from the transducer. Here we refer to F_1 as forward-scattering function. Equations (1) and (2) lead to

$$p_f = p_d \frac{r_d}{r_i r_1} F_1.$$

It is clear from this equation that the phase difference between p_d and p_f is the phase of the forward-scattering function. Further the forward-scattering amplitude can be determined by

$$|F_1| = \frac{P_f}{P_d} \frac{r_1 r_i}{r_d}, \quad (3)$$

where P_f and P_d represent the amplitude. The ranges in Eq. (3) can be accurately measured from the time of arrival of signals and the sound speed in the water. The pressure amplitudes (or voltages) are measured with the aforementioned procedure, and thus we are able to determine $|F_1|$. For the bistatic scattering, we can replace r_1 by r_2 , which is the distance from the target to the second hydrophone and use Eq. (3) to determine the scattering amplitude. However, the phase of the bistatic scattering function cannot be determined by the phase difference between p_d and p_f , since there is path difference between them.

For measurement of backscatter, we use a copper sphere of 60-mm diameter for standard calibration (Foote, 1983; Miyano et al., 1993). The principle is described briefly as follows. Let P_{bs}^0 be the wave amplitude backscattered by the sphere and $|F_2^0|$ be the backscattering amplitude of the sphere. Then, for the sphere

$$P_{bs}^0 = \frac{P_0}{r_0^2} |F_2^0|, \quad (4)$$

where r_0 is the distance from the sphere to the transducer. Similarly for the fish, the received backscattered amplitude is related to the backscattering amplitude $|F_2|$ by

$$P_{bs} = \frac{P_0}{r_i^2} |F_2|, \quad (5)$$

and from Eqs. (4) and (5) we find

$$|F_2| = \frac{P_{bs}}{P_{bs}^0} \frac{r_i^2}{r_0^2} |F_2^0|. \quad (6)$$

Since the backscattering cross section of the sphere $\sigma_{bs}^0 = |F_2^0|^2$ is precisely known, the backscattering amplitude of the fish is determined by Eq. (6). From the scattering function, one can define target strength as

$$TS = 20 \log_{10} \left(\frac{|F_{1,2}|}{R_1} \right), \quad (7)$$

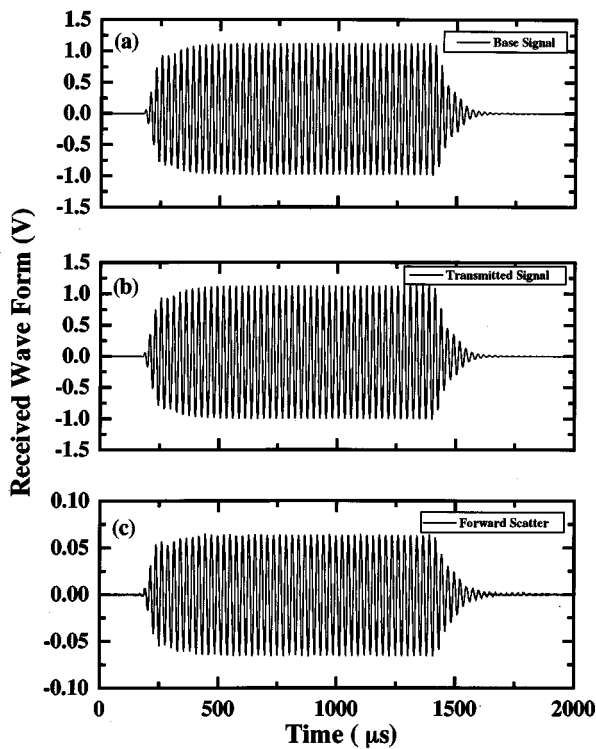


FIG. 2. (a) Signal received at the first hydrophone without any target (base signal). (b) Signal received in the presence of the copper sphere (transmitted signal). (c) Forward-scattered signal obtained by subtracting (a) from (b).

where $R_1=1$ m. In the following presentation of data, the scattering amplitude is given in terms of target strength.

So far we have assumed that the target is on the beam axis. If it deviates slightly, the results can be corrected by dividing Eq. (3) by $D(\theta)$ and Eq. (6) by $D^2(\theta)$, where $D(\theta)$ is the beam pattern of the transducer and θ the position of the target determined with the split beam.

III. RESULTS AND DISCUSSIONS

A. Artificial targets

Two artificial targets were used to test the forward-scattering measurement. One was the copper sphere with a diameter of 60 mm used for standard calibration of back-scattering measurement, and the other was a prolate spheroid made of styrofoam with a major axis of 200 mm and a minor axis of 40 mm. The scattering function of a fixed sphere comprising solid isotropic materials supporting both compressional and shear waves has been well studied and the analytical solution is available (Faran, 1951; Hickling, 1962). Therefore we start first with such a target and then compare with the theory. Figure 2 shows the received waveform at the first hydrophone without the sphere [Fig. 2(a), referred to as base signal hereupon] and with the sphere [Fig. 2(b)]. The sphere was 1.24 m below the hydrophone, and on the beam axis. The ping interval is 613.3 ms and the number of averaging is 150. The sampling rate for the data is 2 MHz. Although Fig. 2(a) and (b) show no visible difference, their difference [Fig. 2(b) minus Fig. 2(a)] is detectable as shown in Fig. 2(c). To determine the wave amplitude, we chose to calculate the standard deviation of the flat part of the signal

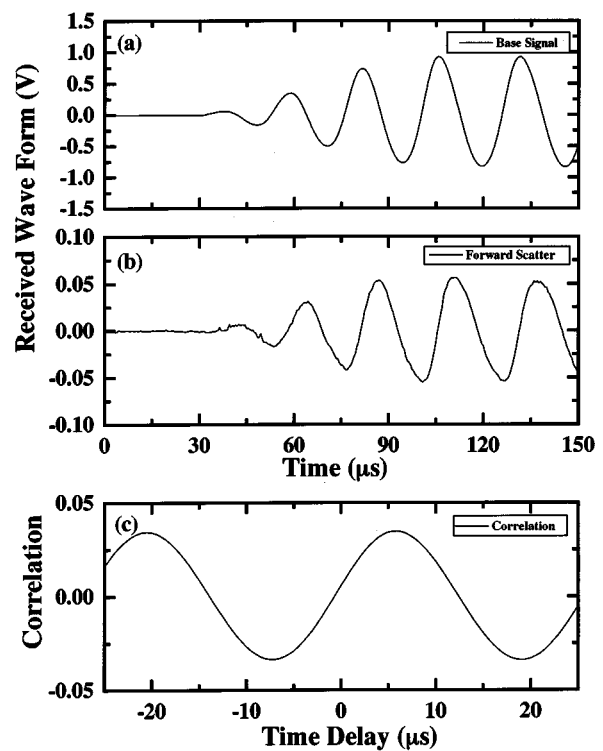


FIG. 3. (a) Expanded part of Fig. 2(a). (b) Expanded part of Fig. 2(c). (c) Cross correlation of (a) and (b).

(e.g., from 400 to 1400 μ s), instead of detecting the peak amplitude from the signal. The standard deviations from Fig. 2(a) and (c) were used in Eq. (3) to determine the forward-scattering amplitude, which was found to be -23.70 dB in terms of target strength defined in Eq. (7). Figure 3(a) and (b) show expanded parts of Fig. 2(a) and (c). It is seen that the forward scattered signal is delayed relative to the base signal. Cross correlation [Fig. 3(c)] was also performed to determine the time delay, which shows the delay is 6μ s, corresponding to a phase shift of -82° . From the above derivation, this phase difference corresponds to the phase of the forward-scattering function. The water temperature during the experiment was 10.5°C , corresponding to a sound speed of 1450 ms^{-1} for fresh water. Theoretical calculation using Hickling's formulas shows that at this sound speed and at frequency 38 kHz, the forward-scattering amplitude is -23.90 dB and the relative phase shift is -55° . Therefore the results were consistent with theory. From both theory and experiment (Miyano et al., 1993), the backscattering target strength of the copper sphere in the same condition is -33.7 dB. It is seen that the forward scattering is 10 dB stronger than the backscattering.¹

Figure 4 shows the same plot as Fig. 2 for the signals received at the second hydrophone, which relative to the sphere is at an angle of 33° off the beam axis. It is seen that Fig. 4(b) has a visibly larger amplitude than Fig. 4(a) and an extended tail. This is due to the scattered signal shown in Fig. 4(c) which arrives later than the direct path signal. The scattering amplitude is estimated from Eq. (3) and found to be -32.33 dB. The phase is difficult to determine since it would require very precise measurement of the path lengths.

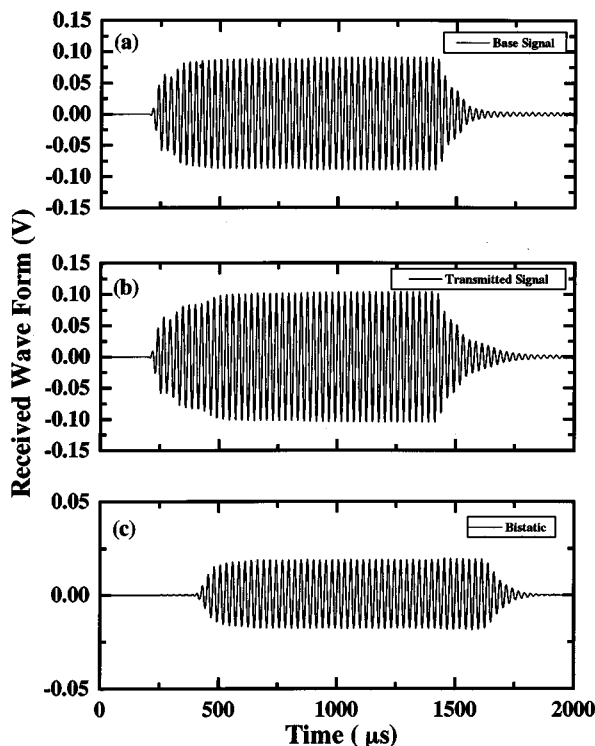


FIG. 4. Bistatic scattered signal from the copper sphere received at the second hydrophone. Captions are the same as in Fig. 2.

From the theoretical calculation, the scattering amplitude at the angle is -31.69 dB.

We repeated such measurements several times, and found that for forward scattering, the results ranged from -24.48 to -22.80 dB. For the bistatic scattering, the results were much more stable, ranging from -31.70 to -32.48 dB, which is expected since the direct-path signal is much weaker in this case (21 dB below the on-axis value). It is also noted that the measured bistatic scattering amplitude is consistently smaller than the theoretical value. We suspect that this may be due to the fact that the scattered wave arrives at an angle (57°) relative to the axis of the hydrophone which is a cylinder, and thus the sensitivity is slightly lower.

For the styrofoam prolate spheroid, we use Stanton's deformed cylinder model (Stanton, 1989) to calculate the forward scattering at normal incidence. We assume that the acoustic properties of styrofoam are nearly equal to those of air. Then we find that the forward-scattering amplitude is -13.2 dB and the backscattering -26.0 dB. A number of independent measurements showed the result ranged between -13.6 and -14.6 dB, and the backscatter was generally 12 dB weaker. Given that the model is approximate and the acoustic properties of styrofoam were not accurately determined, the agreement is very reasonable.

We have also tried different ping rates ($T_p = 346.6$, 613.3 , and 1146.6 ms) and numbers of averages ($N = 50, 100, 150, 200, 300$). The selection criterion is that the difference between two independently measured base signals is minimum, since as described above, the free-field signal in the forward-scattering measurement is assumed unchanged over time. By trial and error, we found that for $T_p = 613.3$ ms, $N \geq 150$ and $T_p = 1146.6$ ms, $N \geq 100$ the re-

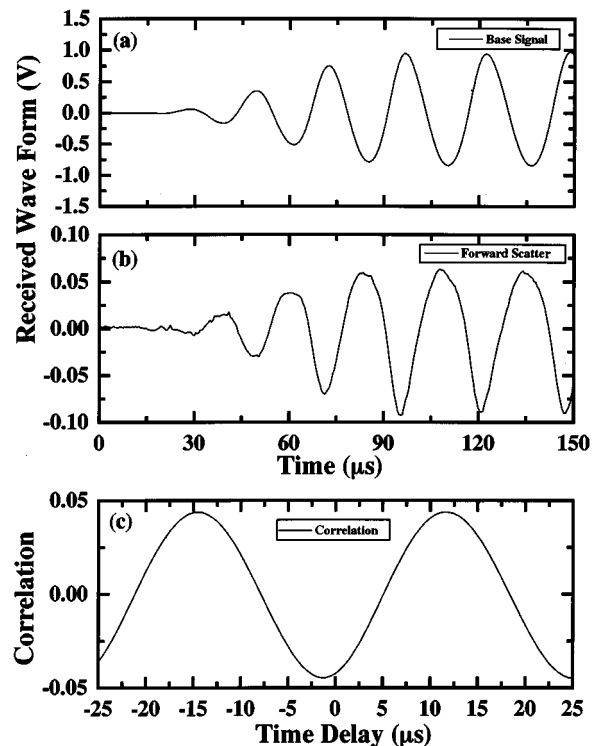


FIG. 5. Free-field signal (base signal), forward-scattered signal by the fish, and their cross correlation. Captions are the same as in Fig. 3.

sults are relatively stable. We then further decided that $T_p = 613.3$ ms and $N = 150$ yielded the most satisfactory result (with the difference less than 1%) while not requiring too long an experimental period. This was used for all the measurements.

B. Fish

We measured four Japanese mackerels (*Scomber Japonicus*) in our initial trial, and the choice of the species was primarily for convenience, rather than for a detailed study of a specific species. In addition, the Japanese mackerel is an important species in the Japanese seas. The fish were rather fresh and x-ray imaging clearly revealed the swimbladders. Figure 5 shows the expanded base signal, the forward-scattered signal, and their cross correlation for a Japanese mackerel with a fork length of 33.4 cm and a weight of 424 g, in the case of normal incidence and in the side aspect. The swimbladder, x rayed immediately before the measurement, had a length of 9.02 cm. The water temperature was between 11.0° and 11.5° C. It is seen that the forward scattered signal is delayed and the phase shift -157° . The forward-scattering amplitude is estimated to be -17.31 dB. The fork length and weight of the fish used for the trial ranged from 31 to 34 cm and from 369 to 452.5 g. The measured forward-scattering amplitude ranged from -19.33 to -16.45 dB in the side aspect and from -21.65 to -17.78 dB in the dorsal aspect. There is a tendency that longer fish result in stronger forward scattering, and the weight appears to have little effect. However, more data are needed to support this observation.

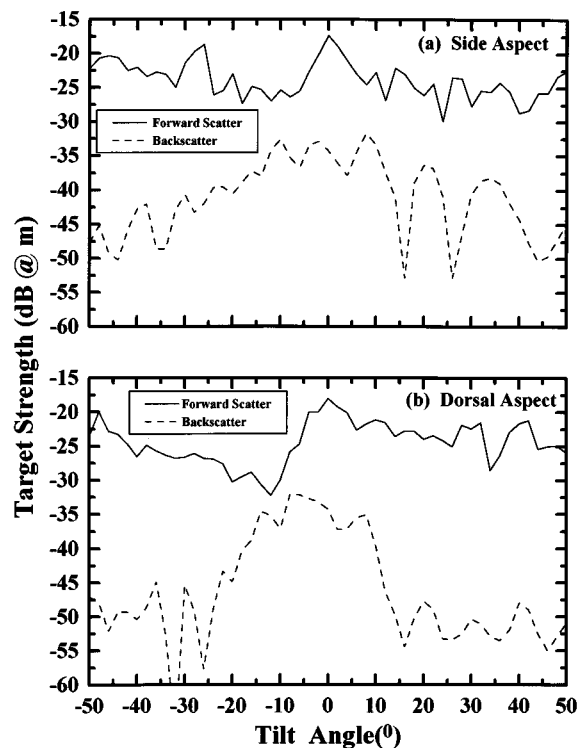


FIG. 6. Dependence of the forward-scattering and backscattering amplitude of the fish on the tilt angle. Solid line for forward scattering and broken line for backscattering. (a) Side aspect. (b) Dorsal aspect.

In measuring the dependence of forward-scattering amplitude on the incident angle of sound, the tilt angle was incremented by 2° , and ranged from -50° to 50° . The base signal was measured three times during each measurement: before placing the fish in the water, in the middle of the measurement (e.g., after the angle had varied from 0° to -50°), and after finishing the measurement. The purpose was to make sure the hydrophone positions remained unchanged thus maintaining a constant phase in the received signal. We intercompared the three measured base signals, discarded any one if it gave anomalous results of forward-scattering amplitude, and then chose the most appropriate one. In addition, we directly monitored the resulting forward-scattered signal on the oscilloscope. If there was sudden change in the result at any tilt angle, we repeated the measurement once. The reason was that once a while, an unknown moving target was observed with the echo sounder. Although the amplitude was small compared to the echo from the fish and the coherent averaging further reduced its effect, it may have had some effect if it happened to be close to the hydrophone during the process of the coherent averaging.

We now plot in Fig. 6 the forward-scattering amplitude in terms of target strength versus the tilt angle for the same fish as in Fig. 5. Simultaneous measurements of the backscattering target strength (coherent summation was also performed) are also shown in the same plots. In these figures, we have corrected for the deviation of the fish from the beam axis, as described earlier. The maximum deviation is 3.4° when the tilt angle is 50° . The 3-dB beam width is 3.1° (one side), and thus part of the fish was beyond the 3-dB beam.

For the tilt angle smaller than 40° , the fish's center was well within the beam. Therefore, except at larger angles, most part of the fish was within the beam.

Perhaps the most obvious feature that can be seen from Fig. 6 is that the forward-scattering amplitude is quite significantly larger than the backscattering amplitude. For example, the maximum value of the forward-scattering amplitude near normal incidence is 14–15 dB higher than that of the backscattering amplitude. The results from the other fish also show the forward scattering 14–17 dB stronger. This is similar to what has been observed from the copper sphere and the styrofoam prolate spheroid. A second feature to be noticed is that the forward scattering varies with the tilt angle to a much less extent than the backscattering. Take Fig. 6(b) as an example. The backscattering at larger angles (say larger than 20°) can drop by more than 25 dB from the peak near normal incidence, while the forward scattering drops by a maximum of 14 dB. The side-aspect result in Fig. 6(a) shows slightly less variability. Measurements with the other fish show similar variability in both forward scattering and backscattering. It is also seen that the forward scattering at larger angles tends to increase after some drop while the backscattering tapers off with increasing angle (more obvious for the negative angles).

It should also be pointed out that due to the off-axis deviation, the forward scattering becomes bistatic, and the results presented here could be somewhat different from what would have been obtained if the fish were kept on the axis. In a separate study, we used a sound scattering model developed in Ye *et al.* (1996) to calculate forward scattering based on the body and swimbladder dimensions of the same fish (Ding and Ye, 1996). Without consideration of the off-axis deviation, the model produces very smooth results. When the deviation is incorporated, however, the model reveals variability similar to that of the data near normal incidence. Therefore, it is reasonable to believe that much of the variability near normal incidence in the data is not inherent but due to the off-axis deviation.

One possible explanation for the less sensitivity of forward scatter could be drawn from the view of interference of scattered signals from different parts of a fish, as suggested by one of the reviewers. Consider a pair of scatterers with spacing d for simplicity. The phase difference between the backscattered signals is proportional to $2d \sin \theta$, where θ is the angle of incidence relative to the line connecting the two scatterers. Depending on the angle and frequency, this phase difference can cause strong interference and thus the total backscattered signal can vary significantly. In the forward direction, however, the path difference (from source to scatterer to receiver) is zero, and thus there is no phase difference between the forward-scattered signals. While fish body and swimbladder are far more complicated, such a simple consideration does provide some insight on the variability of forward scatter and backscatter.

The bistatic-scattering amplitude is shown in Fig. 7 against the tilt angle. Relative to the fish, the receiving hydrophone is at an angle of 21° off the beam axis of the transducer (Fig. 1). When the tilt angle is positive (the fish head moving towards the transducer, and the axis normal to the fish moving farther away from the hydrophone), the scat-

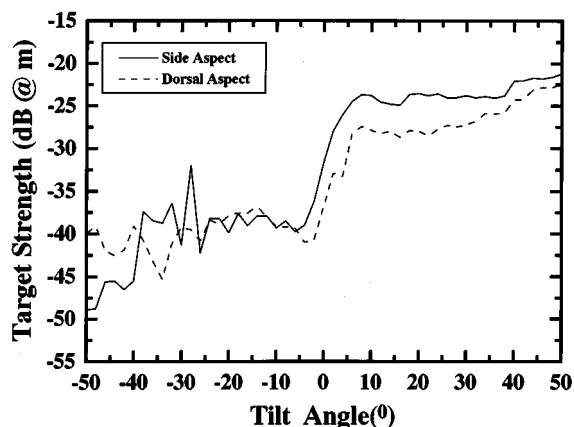


FIG. 7. Dependence of the bistatic scattering amplitude of the fish on the tilt angle. Solid line for the side aspect and broken line for the dorsal aspect.

tering amplitude increases rapidly and then stabilizes at some level. When the angle is negative, the scattering drops first and then stabilizes. Such a feature can be studied using an appropriate theoretical model, but this is beyond the scope of this paper. Bistatic scattering characteristics are important for use of an array of receivers to improve spatial resolution. A better way to study bistatic scattering is to use a hydrophone array, as is presently underway.

IV. CONCLUDING REMARKS

Here we have described a novel laboratory experiment for measurement of forward scattering by individual fish. Simultaneous backscattering measurements have also been provided. The results demonstrate that the forward scattering is significantly stronger than the backscattering, and varies to a less degree with the incident angle of sound. This experiment is an initial trial for testing the measurement concept, and is not aimed at studying a specific species. It is clear that more measurements with various sizes and species of fish and with different frequencies are necessary for more complete understanding of forward scattering. It must be emphasized that similar to backscattering measurement, forward-scattering measurement is also sensitive and that the details of the dependence on incident angle can vary significantly from one fish to another. An in-depth understanding of the results requires more careful studies using appropriate scattering models, which is the subject of another paper (Ding and Ye, 1996).

Finally, it should be pointed out that the time of arrival of forward-scattered signals is essentially independent of the target range and thus the principle cannot be used to measure the target range, unless with an array of receivers. Use of an acoustic array would require a more careful study of bistatic scattering, and would complicate the measurement system. Perhaps simultaneous employment of echo sounders is a simpler solution. Given that the forward scattering is much stronger with less variability, a more challenging and important question is how to make the best use of forward scatter to detect and quantify schools of fish in real applications.

ACKNOWLEDGMENTS

This work was supported by the Science and Technology Agency (STA) of Japan, in the form of a visiting fellowship. The author is indebted to the Fisheries Acoustics group at the National Institute of Fisheries Engineering (NRIFE) for generous support. T. Okumura in the group is thanked for helping with the initial experimental setup. Dr. M. Furusawa of Tokyo University of Fisheries, who invited the author to visit the NRIFE as an STA fellow, is thanked for carefully reviewing the manuscript and providing useful comments and references. Valuable discussions on forward scattering with Dr. Zhen Ye, Dr. Mark Trevorrow, and Dr. David Farmer at the Institute of Ocean Sciences, Canada, are gratefully acknowledged.

¹The difference between forward-scattering and backscattering strengths for a rigid sphere in the Kirchhoff region can also be estimated roughly as follows: $TS_f - TS_b = 20 \log_{10}(k a^2/2) - 20 \log_{10}(a/2)$, where a is the radius and k the wavenumber [see Crispin and Siegel (1960)]. Without going into details of the exact solution, the forward scattering of the copper sphere is estimated to be about 14 dB higher than its backscattering. (Comments from one of the referees.)

- Clay, C. S., and Medwin, H. (1977). *Acoustical Oceanography* (Wiley, New York).
- Crispin, J., and Siegel, K. M. (eds.) (1960). *Methods of Radar Cross-section Analysis* (Academic, New York), pp. 156–159.
- Curran, T., Lemon, D., and Ye, Z. (1994). “The acoustic scintillation flowmeter: Applications for a new environmental tool,” *J. Assoc. Canadian Hydrographic Service*, Spring, 24–29.
- Ding, L., and Ye, Z. (1996). “A method for acoustic scattering by slender bodies. II Comparison with laboratory measurements,” *J. Acoust. Soc. Am.* (submitted).
- Faran, J. J., Jr. (1951). “Sound scattering by solid cylinders and spheres,” *J. Acoust. Soc. Am.* **23**, 338–344.
- Farmer, D. M., and Clifford, S. F. (1986). “Space time acoustic scintillation analysis: A new technique for probing ocean flows,” *J. Ocean Eng.* **OE-11**(1), 42–50.
- Foote, K. G. (1983). “Maintaining precision calibrations with optimal copper spheres,” *J. Acoust. Soc. Am.* **73**, 1054–1063.
- Furusawa, M., Ishii, K., and Miyanoana, Y. (1992). “Attenuation of sound by schooling fish,” *J. Acoust. Soc. Am.* **92**, 987–994.
- Furusawa, M., Takao, Y., Sawada, K., Oukubo, T., and Yamatani, K. (1993). “Versatile echo sounding system using dual beam,” *Nippon Suisan Gakkaishi* **59**, 967–980.
- Hickling, R. (1962). “Analysis of echoes from a solid elastic sphere in water,” *J. Acoust. Soc. Am.* **34**, 1582–1592.
- ICES (1996). *Fisheries and Plankton Acoustics*, edited by E. J. Simmonds and D. N. MacLennan, ICES J. Mar. Sci. (Academic, London), Vol. 53, No. 2.
- Ishimaru, A. (1978). *Wave Propagation and Scattering in Random Media* (Academic, New York), Vol. 2.
- Kimura, K. (1929). “On the detection of fish-groups by an acoustic method,” *J. Imp. Fish. Inst. (Tokyo)* **24**, p. 1141–1152.
- MacLennan, D. N., and Simmonds, E. J. (1992). *Fisheries Acoustics* (Chapman and Hall, London).
- Miyanoana, Y., Ishii, K., and Furusawa, M. (1993). “Spheres to calibrate echo sounders at any frequency,” *Nippon Suisan Gakkaishi* **59**, 933–942.
- Stanton, T. K. (1989). “Sound scattering by cylinders of finite length. III. Deformed cylinders,” *J. Acoust. Soc. Am.* **86**, 691–705.
- Ye, Z. (1995). “Theoretical description of possible detection of swimbladder fish in forward scatter,” *J. Acoust. Soc. Am.* **98**, 2717–2725.
- Ye, Z. (1996). “On acoustic attenuation of swimbladder fish,” *J. Acoust. Soc. Am.* **100**, 669–672.
- Ye, Z., and Farmer, D. M. (1996a). “Acoustic scattering by fish in forward direction,” *ICES J. Mar. Sci.* **53**, 249–252.
- Ye, Z., and Farmer, D. M. (1996b). “Acoustic detection of fish school by reciprocal sounding system: Theory,” *J. Acoust. Soc. Am.* (submitted).
- Ye, Z., Hoskinson, E., Dewey, R. K., Ding, L., and Farmer, D. M. (1996). “A method for acoustic scattering by slender bodies. I. Theory and verification,” *J. Acoust. Soc. Am.* (submitted).

Anomalous attenuation effect on reflectivity of an ultrasonic wave from a thin layer between dissimilar materials

A. I. Lavrentyev and S. I. Rokhlin

The Ohio State University, Nondestructive Evaluation Program, 190 West 19th Avenue, Columbus, Ohio 43210

(Received 30 September 1996; accepted for publication 12 February 1997)

Distinctive features of ultrasonic spectroscopy of adhesive joints of dissimilar materials (including anisotropic) are addressed both in theory and experiment. It is found that for dissimilar joining materials the depth of the reflection spectrum minimum depends nonmonotonically on attenuation in the layer and differs for measurements from opposite sides of the joint. It depends on both the attenuation in the adhesive and on adhesive/substrate impedance mismatches, resulting sometimes in zero reflectivity at nonzero attenuation in the adhesive. The effect is observed at normal and oblique incidence. A technique for measurement of the layer attenuation from amplitude and phase spectra is proposed. Experimental results show good agreement with theoretical predictions. © 1997 Acoustical Society of America. [S0001-4966(97)03706-5]

PACS numbers: 43.35.Cg, 43.35.Ns, 43.35.Zc, 43.60.Pt [HEB]

INTRODUCTION

Ultrasonic spectroscopy of a layer between two materials has been developed, mostly in the 1970s, for nondestructive testing of adhesive joints. For a thin layer separating two substrates the signals reflected from front and back surfaces of the layer overlap in the time domain and interfere. Chang *et al.*¹ and Flynn² used ultrasonic spectroscopy to determine velocity and attenuation and correlated them with the joint cohesive strength. The effect of adhesive attenuation on amplitude and phase spectra of the signal reflected from the joint bondline was studied in Ref. 3. Through-thickness resonance measurements were used in Refs. 4 and 5 to calculate both thickness and modulus of the adhesive layer. Several ultrasonic techniques for evaluation of thin layers have been reported by Kinra *et al.*^{6,7} Normalized amplitude spectra were used in Ref. 8 to measure attenuation. A review of the field is given in Ref. 9. The emphasis in most studies was on the determination of adhesive layer properties in joints of similar materials (Al/Al or steel/steel) by a normally incident ultrasonic wave. For characterization of the adhesion between adhesive and substrate the measurement should be performed in such a way that the ultrasonic wave produces shear stress on the interface¹⁰ as, for example, angle beam ultrasonic spectroscopy.¹¹⁻¹³

Interference of the ultrasonic signals reflected from the front and back of the adhesive layer results in the appearance of minima in the spectrum of the signal reflected from the joint bondline. One would expect from previous experience that an increase of adhesive layer attenuation leads to widening and a depth decrease of the spectral minimum. In this paper we show that this is not always true for a layer between dissimilar materials. In certain conditions attenuation increases results in deepening and narrowing of the reflection minimum. A preliminary discussion of this phenomenon has been given in Ref. 15. In the following sections the phenomenon is studied theoretically and experimentally.

I. ANOMALOUS ATTENUATION EFFECT

Let us consider an ultrasonic plane wave normally reflected from a layer separating two similar substrates (for example, an adhesive layer in an adhesive joint between aluminum plates). Usually the bonded plates are thick enough that the signals reflected from the plate surfaces are separated in time from the signals reflected from the adhesive layer. Ultrasonic pulses with a common center frequency of 10 MHz and a bandwidth at -10 dB of 10 MHz reflected from the front and back sides of the adhesive layer are not separated in time and interfere (a usual adhesive layer thickness is about $100\text{ }\mu\text{m}$). Gating the signals reflected from the layer one finds that the frequency spectrum of the signal has minima which are due to destructive interference of the signals reflected from the front and back sides of the layer. The first minimum corresponds to the half wavelength resonance, the second to the one and a half wavelength resonance and so

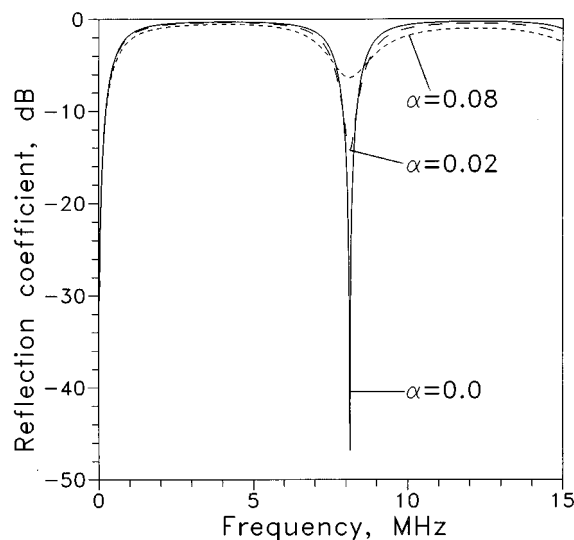


FIG. 1. Spectra of the ultrasonic signal reflected from an adhesive layer between two aluminum substrates at normal incidence.

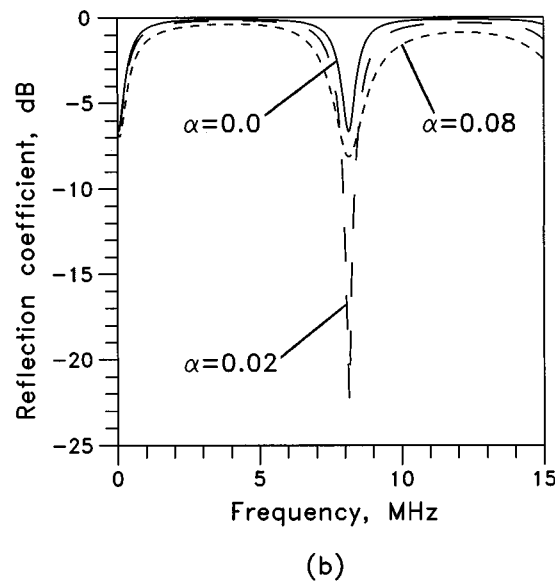
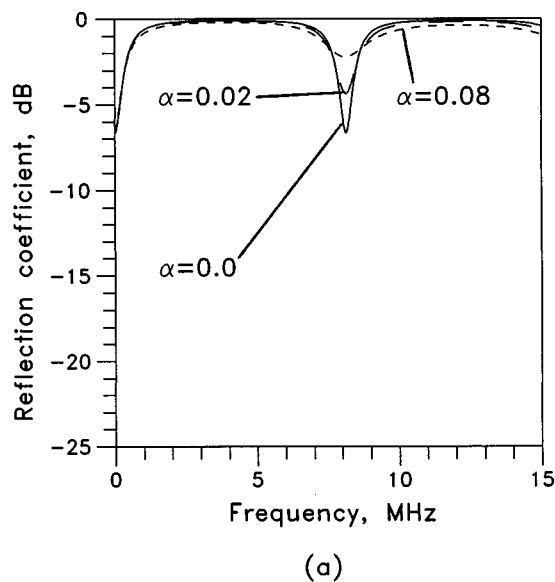


FIG. 2. Spectra of the ultrasonic signal reflected from an adhesive layer between aluminum and steel substrates at normal incidence: (a) incidence from steel side, (b) incidence from aluminum side.

on. Figure 1 shows calculated spectra for different attenuations in the layer between Al substrates. For calculation the thickness and properties of the layer are taken to be close to those of adhesive met in practice: $h_{\text{adh}} = 135.8 \mu\text{m}$, $Z_{\text{adh}} = 2.5 \times 10^6 \text{ (g/cm}^2 \text{ s)}$. At zero attenuation ($\alpha = k''/k' = 0$) the reflectivity at minimum equals zero. Increase of attenuation results in nonzero reflectivity at the minimum, with higher attenuation corresponding to greater reflection coefficient. Attenuation can be uniquely determined from the depth and the width of the spectral minimum.

Figure 2(a) and (b) show calculated spectra for an ultrasonic wave reflected from an adhesive layer between aluminum and steel substrates [(a), incidence from steel side and (b), from aluminum side]. The thickness and properties of

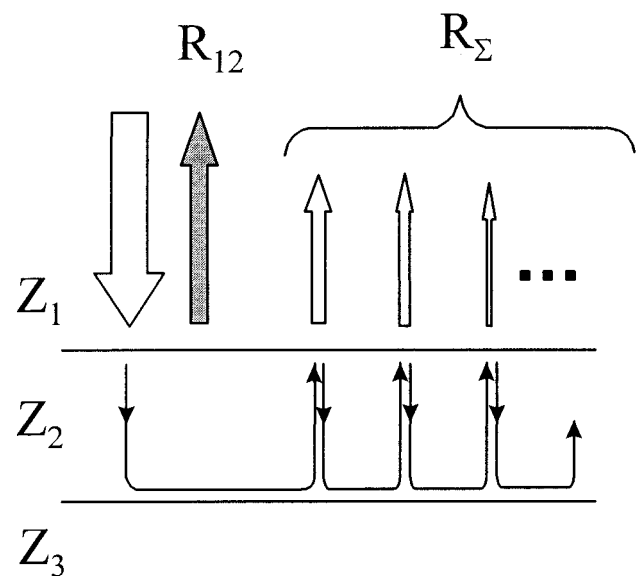


FIG. 3. Schematic of ultrasonic wave reflection from a layer enclosed between two different materials.

the adhesive layer are the same as in the previous case. At zero attenuation the spectra obtained from aluminum and steel sides of the joint are identical (solid lines). In both cases reflectivities at minimum equal that at the aluminum/steel interface, e.g., at half wavelength resonance the adhesive layer is transparent to ultrasonic waves. Increase of attenuation leads to increase of reflectivity measured from the steel side, as for identical substrates. For incidence from the aluminum side increase of attenuation initially leads to decrease of reflectivity (deepening of the reflection minimum). At “critical” attenuation ($\alpha = k''/k' \approx 0.029$) the reflectivity vanishes; further increase of attenuation results in reflectivity increase.

To understand this phenomenon consider an ultrasonic wave reflecting from an adhesive layer enclosed between two substrates (Fig. 3). The incident signal has unit amplitude. The total reflection from the layer can be separated into two interfering signals: (a) reflection from the top surface of the layer (first-reflection signal) with amplitude R_{12} and (b) reflection from the bottom which is the sum of all possible multiple reflections inside the layer R_{Σ} (multiple-reflection signal). At resonance, when the phases of these two signals are opposite, they partially cancel each other, with the resulting signal amplitude depending on the ratio of the interfering signal amplitudes.

For identical substrates ($Z_1 = Z_3$) at zero attenuation the amplitudes of the first-reflection and multiple-reflection signals are equal and therefore the reflectivity of the layer is zero (Fig. 1, $\alpha = 0$). Nonzero attenuation has negligible influence on the first-reflection signal but decreases the amplitude of the multiple-reflection signal. As a result the amplitude cancellation between first- and multiple-reflection signals at the resonance is not complete and the layer has nonzero reflectivity, with higher attenuation giving greater reflectivity (Fig. 1, $\alpha > 0$).

When the upper substrate impedance is higher than that of the lower substrate and the attenuation in the layer is zero

the multiple-reflection signal amplitude is smaller than that of the first-reflection, thus at the resonance the total reflectivity of the layer is not zero [Fig. 2(a), $\alpha=0$]. Increase of attenuation in the layer further decreases the multiple-reflection signal amplitude; thus minimum depth decreases with attenuation.

When the upper substrate impedance is smaller than that of the lower and the attenuation in the layer is zero the amplitude of the first-reflection signal is smaller than that of the multiple-reflection signal. At the resonance, increase of attenuation decreases the amplitude of the multiple-reflection signal (the first-reflection signal is almost unaffected) thus decreasing the amplitude of the reflected interference signal. At critical (nonzero) attenuation the amplitudes of the first- and multiple-reflection signals are equal and the layer reflectivity equals zero [Fig. 2(b)]. Further increase of attenuation leads to further decrease of the multiple-reflection signal amplitude with consequent decrease of the minimum depth.

II. MINIMUM CONDITIONS AND CRITICAL ATTENUATION

A. Reflection from a layer between dissimilar materials: Exact and approximate solutions

A plane harmonic wave of amplitude A propagating in an attenuating medium may be represented by

$$u(x,t) = Ae^{i(kx - \omega t)}, \quad (1)$$

where u represents displacement, ω represents angular frequency, and the complex wave number $k = k' + ik'' = k'(1 + i\alpha)$, where $\alpha = k''/k'$. k'' is attenuation per unit length and α is attenuation per wavelength. k'' and α are positive so the wave amplitude [Eq. (1)] decreases with propagation distance.

Consider a layer between two different materials as shown in Fig. 3. For incidence from side i the total reflection from the layer can be represented¹⁴ as a sum of the first- and multiple-reflection signals:

$$R_i = R_{i2} - \frac{T_{i2}T_{2i}R_{j2}e^{2ikh}}{1 - R_{12}R_{32}e^{2ikh}}, \quad i, j = 1, 3, \quad i \neq j, \quad (2)$$

where R_{ij} and T_{ij} are complex stress reflection and transmission coefficients. Taking into account that $T_{i2}T_{2i} = 1 - R_{i2}^2$, the layer reflection coefficient can be rewritten as

$$R_i = \frac{R_{i2} - R_{j2}e^{2ikh}}{1 - R_{12}R_{32}e^{2ikh}}, \quad i, j = 1, 3, \quad i \neq j. \quad (3)$$

The complex stress reflection coefficients R_{12} , R_{32} can be represented as

$$R_{j2} = |R_{j2}|e^{i\phi_{j2}}, \quad j = 1, 3, \quad (4)$$

where

$$|R_{j2}| = \sqrt{\frac{(R_{j2}^0)^2 + (\alpha^2/4)(T_{2j}^0)^2}{1 + (\alpha^2/4)(T_{2j}^0)^2}}, \quad (5)$$

$$\phi_{j2} = \arctan\left(-\frac{\alpha}{2} \frac{(1 - (R_{j2}^0)^2)}{(R_{j2}^0 - (\alpha^2/4)(T_{2j}^0)^2)}\right), \quad j = 1, 3, \quad (6)$$

and $R_{j2}^0 = (Z_2 - Z_j)/(Z_2 + Z_j)$, $T_{2j}^0 = 2Z_j/(Z_2 + Z_j)$, $j = 1, 3$ are stress reflection and transmission coefficients and $Z_j^0 = \rho v_j$ is impedance for the media without attenuation. Substituting (4) in (2), we obtain for the layer reflection

$$R_i = \frac{|R_{i2}|e^{i\phi_{i2}} - |R_{j2}|e^{i\phi_{j2}}e^{2ik'h}e^{-2k''h}}{1 - |R_{12}||R_{32}|e^{i(\phi_{12} + \phi_{32})}e^{2ik'h}e^{-2k''h}}, \quad i, j = 1, 3, \quad i \neq j. \quad (7)$$

For a given material system R_{ij}^0 and T_{ij}^0 are constants and the spectral minima position, depth and width depend on k' and k'' .

For a given attenuation k'' the layer reflection minimum occurs when both terms in the numerator of (7) have the same phase (the effect of the denominator is negligible in the vicinity of critical attenuation):

$$(k'h)_i^{(R)} = \pi n + \frac{1}{2}(\phi_{i2} - \phi_{j2}), \quad i, j = 1, 3, \quad i \neq j, \quad n = 1, 2, 3, \dots \quad (8)$$

As discussed in the previous section, at the resonance the first- and multiple-reflection signals (Fig. 3) interfere destructively. For identical substrates ($Z_1 = Z_3$) $\phi_{12} = \phi_{32}$ [Eq. (6)]; consequently, $(k'h)^{(R)} = \pi n$ [Eq. (8)] and minimum reflection frequency does not depend on attenuation. In a joint between dissimilar materials ($Z_1 \neq Z_3$) the reflection minima are shifted from the integer half-wavelength condition by $(\phi_{i2} - \phi_{j2})/2$.

At resonance [Eq. (8)], the layer reflection equals zero when both terms in the numerator of (7) have the same amplitude. This occurs at critical attenuation given by

$$k'' = \frac{1}{2h} \ln\left(\frac{|R_{32}|}{|R_{12}|}\right), \quad n = 1, 2, \dots \quad (9)$$

Since k'' is positive, critical attenuation occurs only when $|R_{32}| > |R_{12}|$ ($Z_3 > Z_1$), i.e., when the wave impinges on the layer from the side of the lower impedance substrate. When $Z_3 > Z_1$ the minimum reflection is greater than zero for any k'' .

The nonlinear equations (8) and (9) are coupled and cannot be solved analytically. To estimate critical attenuation and frequency of minimum reflection simple zero- and first-order approximations can be used. Complex reflection coefficients R_{12} and R_{32} can be expanded in a Taylor series with α as a parameter. Since α in polymer materials is less than 0.1 the higher-order terms can be neglected. In zero-order approximation the dependence of R_{12} and R_{32} on attenuation can be neglected and

$$(k'h)^{(R)} \approx \pi n, \quad n = 1, 2, \dots, \quad (10)$$

i.e., in zero-order approximation the minimum positions do not depend on attenuation. The condition for critical attenuation (9) becomes

$$k'' \approx \frac{1}{2h} \ln\left(\frac{|R_{32}^0|}{|R_{12}^0|}\right), \quad n = 1, 2, \dots \quad (11)$$

In the first-order approximation the linear terms in the Taylor series expansion of R_{j2} , $j = 1, 3$ are kept and we have

$$R_{j2} \approx R_{j2}^0 - i \frac{\alpha}{2} [1 - (R_{j2}^0)^2] \approx R_{j2}^0 e^{i\phi_{j2}^{(1)}}, \quad j=1,3, \quad (12)$$

where

$$\phi_{j2}^{(1)} = -\frac{\alpha}{2} \left(\frac{1}{R_{j2}^0} - R_{j2}^0 \right), \quad j=1,3. \quad (13)$$

The condition for minimum reflection is given by

$$(k'h)_i^{(R)} = \pi n + \frac{\alpha}{4} \xi_i, \quad n=1,2,\dots, \quad (14)$$

where

$$\xi_i = (R_{i2}^0 - R_{j2}^0) \left(1 + \frac{1}{R_{12}^0 R_{32}^0} \right), \quad i,j=1,3, \quad i \neq j. \quad (15)$$

The first-order approximation takes into account the shift of the resonance due to attenuation in the layer. The equation for critical attenuation is the same as for the zero-order approximation given by (11).

B. Effect of substrate impedances on the reflection minimum position and critical attenuation

As follows from (8) and (9) that the minimum reflection position and critical attenuation depend on the dissimilarity of the substrates (Z_3/Z_1). This is illustrated in Fig. 4(a) and (b). The calculation is performed for a sandwich structure of aluminum [$Z_1 = 17.037 \times 10^6 \text{ kg/(m}^2 \text{ s)}$], a polymer layer [$Z_2 = 2.400 \times 10^6 \text{ kg/(m}^2 \text{ s)}$] and a lower substrate with varying impedance Z_3 . Since for polymer materials k'' is approximately proportional (in a limited frequency range) to frequency^{16,17} we assume for numerical simulation $\alpha = k''/k' = \text{const}$. The case of an Al/polymer/steel joint is indicated in the graphs by a vertical dashed line. One can see [Fig. 4(a)] that the reflection minimum shifts to higher frequencies with an increase of Z_3/Z_1 . While the first-order approximation for the minimum position is practically indistinguishable from the exact solution, the zero-order approximation is constant and deviates from the exact value at $Z_3/Z_1 \neq 1$ (for an Al/steel joint the deviation is about 0.001π). Figure 4(b) demonstrates that critical attenuation increases with Z_3/Z_1 and the zero-order approximation is indistinguishable from the exact value.

III. DETERMINATION OF THE ADHESIVE LAYER ATTENUATION

The effect of the polymer layer attenuation on the amplitude of reflection at the resonance in an Al/polymer/steel sandwich is compared in Fig. 5 to that of an Al/polymer/Al system. Reflection amplitudes from aluminum (solid line) and steel (dashed line) sides are shown. The calculation is performed for $k'h = \pi$. From the steel side the reflection coefficient gradually increases with attenuation; it is much higher than that from the aluminum side (when $\alpha > 0$) and its dependence on attenuation is weak and thus the minimum depth measured from the steel side is insensitive to the attenuation. For measurement from the aluminum side the dependence of the resonance reflectivity on the layer attenuation is much greater with $R=0$ at critical attenuation. The

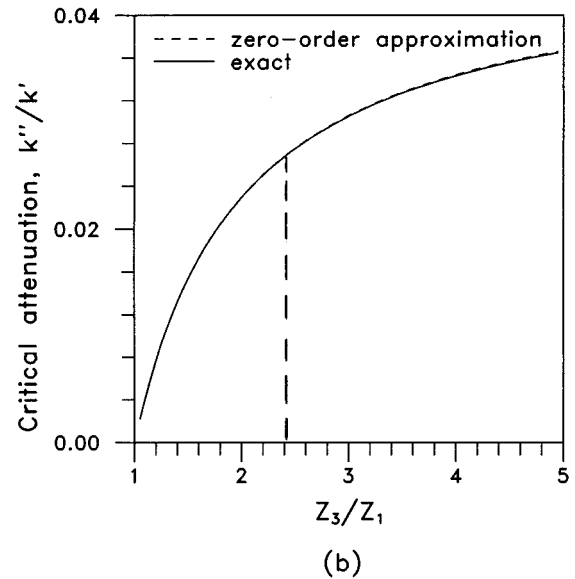
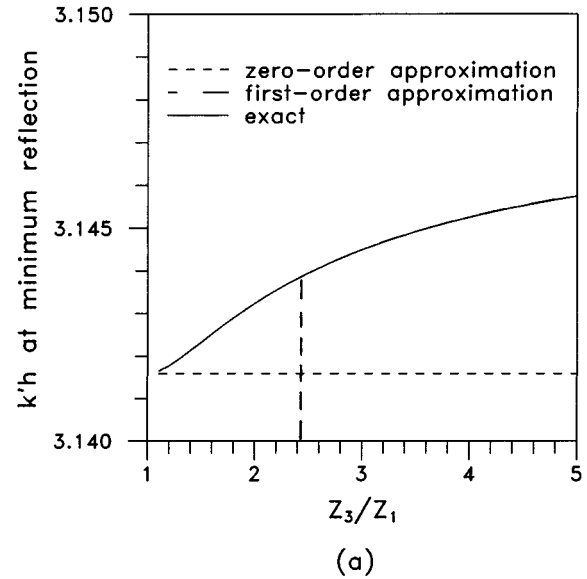


FIG. 4. Effect of impedance mismatch between substrates (Z_3/Z_1) on (a) reflection minimum position and (b) critical attenuation. Calculations are performed in zero- and first-order approximations and exactly.

precision of the minimum depth measurement also depends on the experimental system and sample quality as described in the experimental part of the paper. The reflection minimum depth in the Al/polymer/Al system is shown by the short-dashed line. For attenuations usual for polymer materials (>0.015) the spectral minimum is deeper for an Al/polymer/steel sandwich than for an Al/polymer/Al.

An important characteristic of the spectral minimum reflection amplitude for the system with dissimilar substrates is nonuniqueness of the minima depth on attenuation (when the measurement is performed from the side of the lower impedance substrate). Thus additional information is needed to determine the attenuation from the measured reflection spectrum. The minimum width can help to determine the

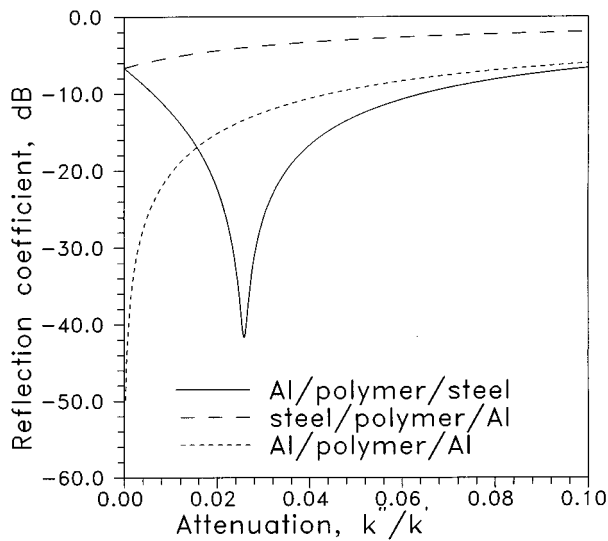


FIG. 5. Dependence of the reflection minimum depth measured in Al/polymer/steel and Al/polymer/Al sandwich on attenuation. Calculations are performed for the resonance position $k'h = \pi$ (zero-order approximation).

attenuation uniquely. Figure 6 shows the dependence of the spectral minimum width on the layer attenuation measured at -3 dB level from the minimum (Fig. 10). The minimum width equals zero at critical attenuation. Dashed lines on the graph connect points corresponding to the same minimum depth; for the same reflectivity higher attenuation corresponds to a greater minimum width. The difference in corresponding widths is significant for a large difference between attenuations and it is negligible when both attenuations are near the critical value (and, consequently, near each other). Thus the minimum width can help determine the attenuation of the polymer layer if the two attenuations giving the same reflectivity are not close.

When the two attenuations are close the minima widths cannot be used as a discriminating parameter. In this case phase spectra can be used as shown in Fig. 7 for the ultra-

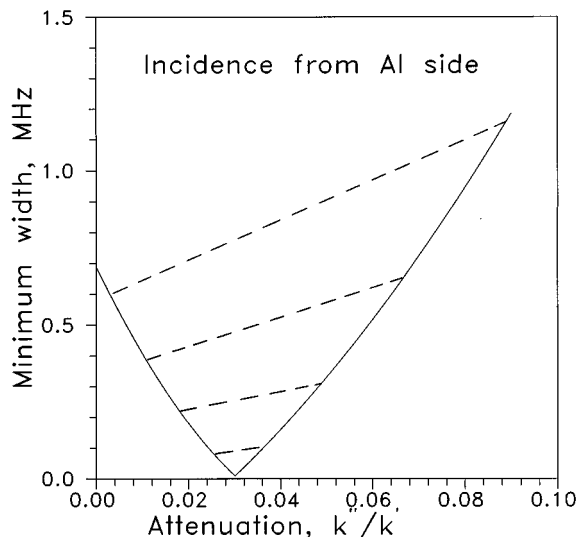


FIG. 6. Dependence of the reflection minimum width on attenuation in the polymer layer.

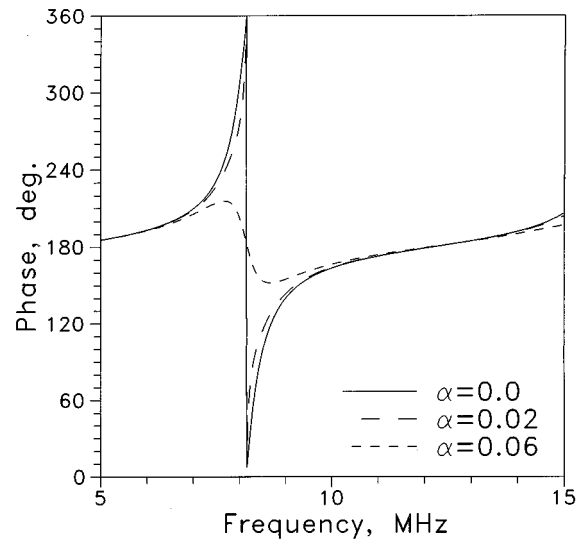


FIG. 7. Phase spectra of the ultrasonic signal reflected at normal incidence from the polymer layer in the Al/polymer/steel sandwich (from the aluminum side).

sonic signal normally reflected from the polymer layer (aluminum side). The critical attenuation α is 0.028 74. One can see that in the vicinity of the resonance the phase is highly dependent on the attenuation, and phase spectra for attenuations on different sides of the critical value can easily be distinguished even if the attenuations are close to each other.

Examples of adhesive layer attenuation determination from reflection minimum depth, width, and phase spectra will be given in the experimental section.

IV. SIGNAL TRANSMISSION AND ENERGY LOSS IN THE LAYER

Measurement of transmission through the polymer layer is an alternative way to measure the attenuation in the layer. Transmissions through layer 2 from material 1 to material 3 (Fig. 3) and from material 3 to material 1 are given by

$$T_{i \rightarrow j} = \frac{T_{i2}T_{2j}e^{ikh}}{1 - R_{12}R_{32}e^{2ikh}}, \quad i, j = 1, 3. \quad (16)$$

Figure 8 shows calculated amplitude transmission spectra for a normally incident longitudinal wave from aluminum to steel through a polymer layer. The transmission in the opposite direction has the same shape in the logarithmic scale since it is defined by the denominator of Eq. (16); it is shifted vertically due to the difference in the numerator. Physically the maximum transmission occurs when the waves reverberating inside the layer interfere constructively. Mathematically, it corresponds to the minimum of the denominator in (16):

$$(k'h)^{(T)} = \pi n - \frac{1}{2}(\phi_{12} + \phi_{32}), \quad n = 1, 2, \dots, \quad (17)$$

where ϕ_{12} and ϕ_{32} are given by Eq. (6). The maximum transmission frequency shifts with increasing α .

Referring to Eq. (2) for the reflection coefficient, the maximum for the second term R_2 occurs at the same condition (17). The minimum in the layer reflection occurs due to destructive interference between first-reflection R_{12} and

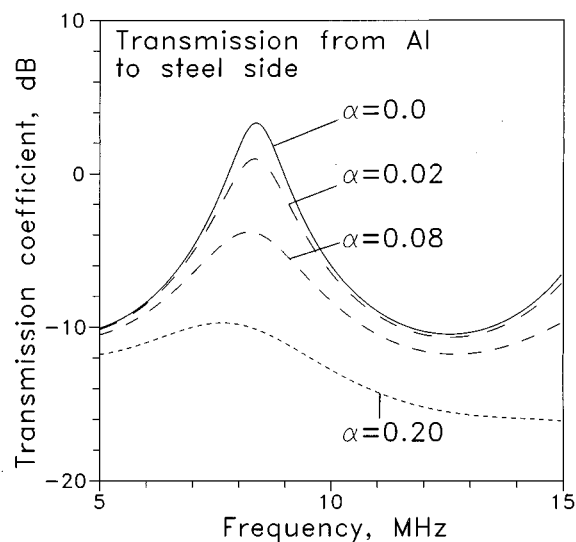


FIG. 8. Amplitude transmission spectra through a polymer layer enclosed between Al and steel plates.

multiple-reflection R_{Σ} signals [first and second terms in (2)]. This destructive interference is more pronounced near the critical attenuation: The denominator in (3) and (16) do not reach zero (if $k'' \ll k'$), while the numerator in (2) equals zero at or in the vicinity of critical attenuation. When the layer attenuation is far from critical the reflection minimum position shifts from (8) toward (17). In the vicinity of the critical attenuation the difference between $(k'h)_i^{(R)}$ and $(k'h)^{(T)}$ [given by Eqs. (8) and (17), respectively] equals ϕ_{12} or ϕ_{32} depending upon which side the wave is incident from.

The attenuation in the layer should be directly related to the energy loss; the higher the attenuation, the greater the loss of energy in the polymer layer. In the plane-wave approximation the energy loss in the layer is given by

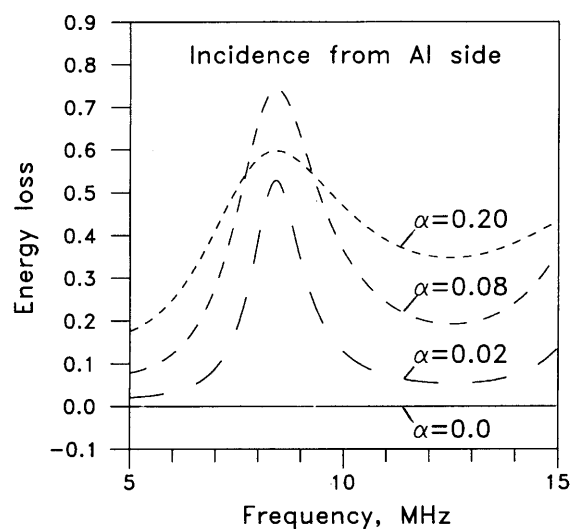
$$E_i = 1 - |R_i|^2 - |T_{1 \rightarrow 3} \cdot T_{3 \rightarrow 1}|. \quad (18)$$

Figure 9(a) and (b) shows energy loss spectra calculated for a longitudinal ultrasonic wave incident normally on the adhesive layer in an Al/steel adhesive joint. The most energy is stored in the resonator (represented by the polymer layer) by the ultrasonic wave at the resonance frequency; and thus the greatest energy loss occurs near the resonance with the frequency of the maximum loss weakly depending on attenuation. The energy loss is smaller when the wave is incident from the steel side [Fig. 9(b)] due to higher steel/polymer impedance mismatch and less energy penetration into the layer. For very high attenuations the energy loss decreases at resonance due to a decrease of the vibration energy while increasing at the other frequencies. In the limit $\alpha \rightarrow \infty$ the energy loss does not depend on frequency and equals $E_i = 1 - |R_{i2}|^2 \rightarrow 0$ since no energy penetrates the layer.

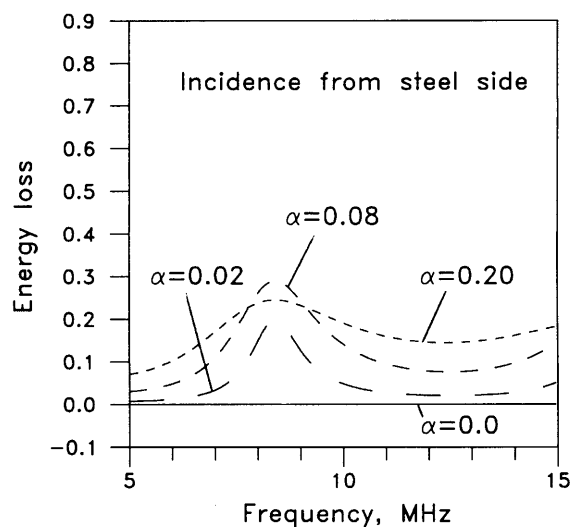
V. EXPERIMENT

A. Experimental approach

Experiments were performed at normal and oblique incidence using an ultrasonic goniometer described in Refs. 12



(a)



(b)

FIG. 9. Energy loss in an adhesive layer: (a) incidence from Al side, (b) incidence from steel side.

and 18. The goniometer allows the measuring of signals reflected from the layer inside the sandwich at various incident angles using only one transducer. Broadband ultrasonic transducers (diameter 0.5 in.) with a central frequency of 10 MHz were used. Before measurements the transducer was aligned perpendicularly to the specimen by obtaining the maximum reflection from the sample. The precision of alignment for a 10-MHz transducer was about $\pm 0.05^\circ$. The ultrasonic signals were amplified, digitized, averaged by a HP 54504A 400-MHz digital oscilloscope, and collected by a computer through an IEEE-488 interface. The data is then processed in the frequency domain and deconvolved with a reference signal reflected from the metal/water interface. The rotation of the transducer, i.e., change of incident angle, and position of the sample using a translation table are controlled

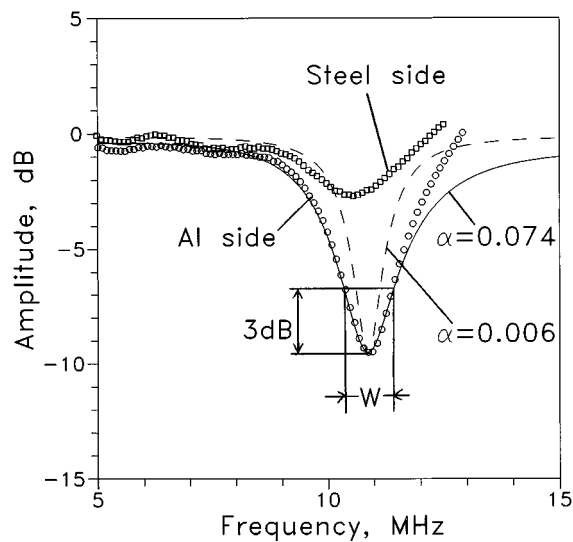


FIG. 10. Experimentally measured and calculated amplitude spectra of the ultrasonic signal reflected normally from the adhesive layer in the Al/steel adhesive joint.

by a computer through an RS-232 interface. Resolution for the rotation table is 0.01° and for the translation table it is 0.01 mm. The apparatus was immersed in water with the temperature stabilized at $29.8 \pm 0.02^\circ\text{C}$ using a Fisher Isotemp Circulator (model 730).

Measurements were performed at normal incidence on different types of joints. Two types of polymer layers were used: (1) FM-73 epoxy resin prepreg manufactured by American Cyanamid, and (2) lexan polymer film. FM-73 prepreg film was used to prepare Al/Al and Al/steel adhesive joints. Adherend surface pretreatment and curing were performed according to aircraft specifications with resulting welded adhesive/adherend interfaces and adhesive layer thickness $h \approx 110\ \mu\text{m}$. The adhesive layer thickness inhomogeneity was less than $3\ \mu\text{m}$.

Lexan polymer film was pressed between polished aluminum and steel plates with glycerin as a couplant with a resulting slip boundary at the metal/polymer interface. The outer edge of the resulting sandwich was sealed. A longitudinal wave at normal incidence interacts with the slip interface in the same way as with a welded interface in the adhesive joint.

In the following discussion we refer to the Al/FM-73 adhesive/steel joint as adhesive joint and to the Al/lexan film/steel sandwich as lexan joint.

B. Amplitude and phase spectra of the reflected ultrasonic signal

Figure 10 shows amplitude spectra of a longitudinal ultrasonic wave reflected normally from the adhesive layer in the Al/steel adhesive joint. Circles correspond to measurement from the Al side, squares to measurement from the steel side. While the positions of the spectral minima are very close to each other the depth observed from the steel side is much smaller than that from the aluminum side. As already discussed the depth of the minimum measured from the steel side uniquely defines the attenuation, but the mini-

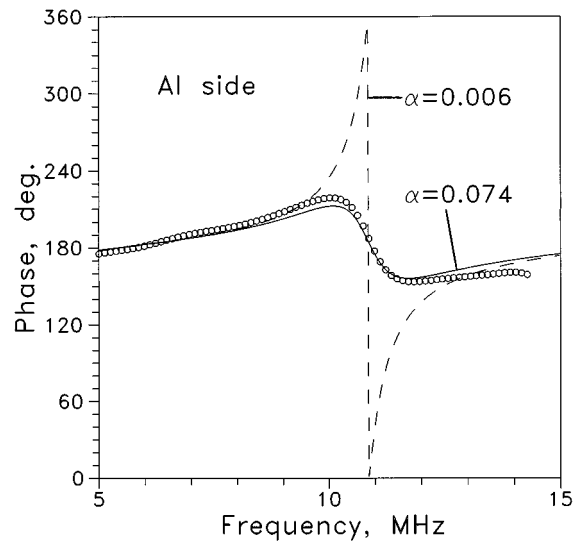


FIG. 11. Experimentally measured and calculated phase spectra of the ultrasonic signal reflected normally from the adhesive layer in the Al/steel adhesive joint from aluminum side.

um depth dependence on attenuation is weak (Fig. 5). The dependence of the minimum depth measured from the aluminum side on the polymer layer attenuation is much greater and can yield precise values of the attenuation. The difficulty is that two different attenuation values give the same minimum depth.

Figure 10 also shows two calculated spectra with the same minima depths as measured experimentally. In this particular case the two attenuations yielding identical minimum depth are very different ($\alpha_1 = 0.006$ and $\alpha_2 = 0.074$), which results in significantly different minima widths (Fig. 6). Since the solid line in Fig. 10 fits the experimental data much better than the dashed line, $\alpha = 0.074$ is the attenuation of the adhesive layer.

Figure 11 shows experimentally measured and calculated phase spectra corresponding to the amplitude spectra shown in Fig. 10. Again, the solid line (corresponding to $\alpha = 0.074$) fits the experimental data much better than the dashed line ($\alpha = 0.006$).

It is more difficult to discriminate between two attenuations yielding the same spectral minimum depth if their values are close to the critical and, thus, to each other. Figure 12 shows spectra (experimental and calculated) of a normally incident longitudinal wave reflected from the aluminum side of the polymer film in the lexan joint. Calculated spectra correspond to attenuations in the polymer layer of 0.021 (dashed line) and 0.031 (solid line). The discrimination can be done using the phase spectra (experimental and calculated) shown in Fig. 13. The attenuation of the polymer film is $\alpha = 0.031$ since the phase spectrum calculated using this value (solid line) fits the experimental data better.

C. Energy loss

Energy loss inside the layer is directly related to its attenuation. Figure 14 shows by circles the frequency dependence of the energy loss calculated using Eq. (18) from the experimentally measured reflected and double-transmitted

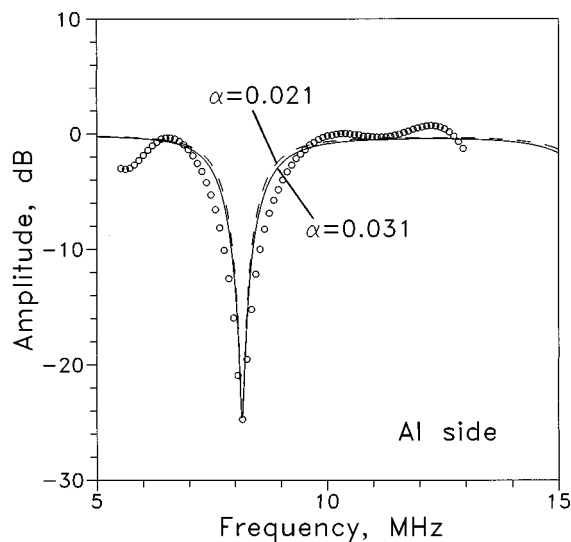


FIG. 12. Experimentally measured and calculated amplitude spectra of the ultrasonic signal reflected normally from the lexan layer in the Al/lexan/steel joint from the aluminum side.

spectra. The theoretical curves for the two attenuations giving the same minimum depth in the reflectivity spectrum are also shown. The solid line (corresponding to $\alpha=0.074$) fits the experimentally measured energy loss, confirming the validity of the determination of the layer attenuation.

VI. OBLIQUE INCIDENCE

A. Theory

Anomalous deepening of the reflected wave spectral minima with increasing attenuation is observed also at oblique incidence. Interpretation of the results in this case is more complex since both longitudinal and transverse waves propagate in the adhesive layer. To demonstrate the phenomenon we performed calculations using a matrix algorithm described in detail in Refs. 19 and 20. For the calculations

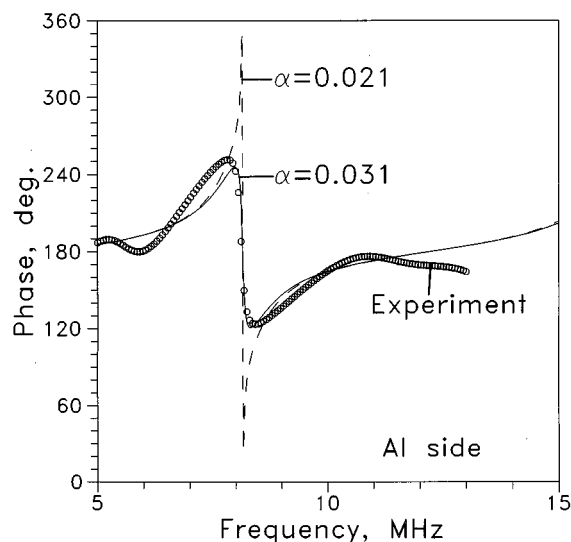


FIG. 13. Experimentally measured and calculated phase spectra of the ultrasonic signal reflected normally from the lexan layer in the Al/lexan/steel joint from the aluminum side.

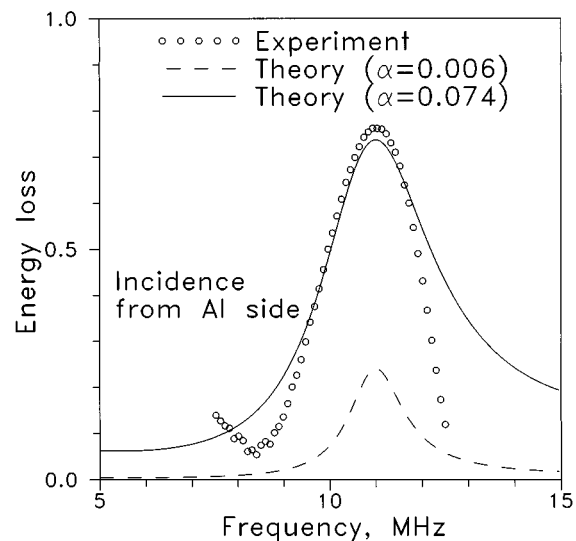


FIG. 14. The frequency dependence of the energy loss calculated from the experimentally measured spectra of the reflected and double transmitted signals and theoretical curves for the two attenuations giving the same resonance reflectivity.

per wavelength attenuations of the transverse and longitudinal waves were assumed the same: $k_t''/k_t' = k_l''/k_l'$ (attenuation per unit length is about twice as great for the transverse wave as for the longitudinal wave). (Alternatively one may use an algorithm described in Ref. 21.)

Figure 15(a) and (b) shows calculated spectra of the oblique transverse ultrasonic (SV) waves reflected from the aluminum side of the adhesive layer inside an Al/steel adhesive joint, for angles of incidence on the Al/adhesive interface of 30° and 60° , respectively. To excite ultrasonic waves in the joint immersed in water angles of incidence on the water/Al interface (14° and 24.8°) greater than the first critical angle of 13° are used, thus only SV waves penetrate the aluminum plate. Both graphs show that attenuation increase from zero to the critical value results in deeper spectral minima, while further attenuation increase results in decreasing minima depths. The critical attenuations for oblique incidence are functions of angle and are not defined by Eq. (9). They can be found numerically.

For an ultrasonic SV wave incident on the Al/adhesive interface at an angle slightly greater than the critical angle at which the reflected longitudinal wave vanishes (for example 30°) the amplitude of the longitudinal wave transmitted into the adhesive is much smaller than that of the transmitted transverse wave (Fig. 16). Thus at this angle of incidence the minima observed in the spectrum of the reflected signal [Fig. 15(a)] are due to interference of an SV wave reflected from the Al/adhesive interface (first-reflection signal) and a multiple-reflection signal composed almost exclusively of an SV wave propagating inside the adhesive layer. Also, the phases of the first- and multiple-reflection signals remain almost opposite. As a result the minima are equally spaced in frequency [Fig. 15(a)]. At 30° the reflection from the Al/adhesive interface (R_{tt}) is smaller than transmission into the adhesive (T_{tt}) and consequently the deepest minimum is observed at nonzero (critical) attenuation. The critical attenua-

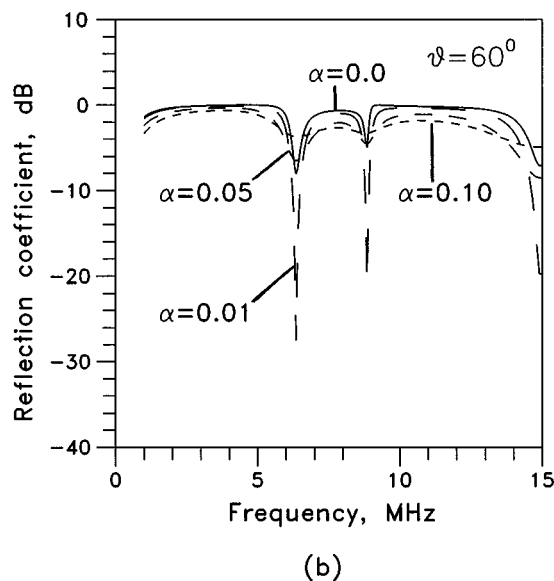
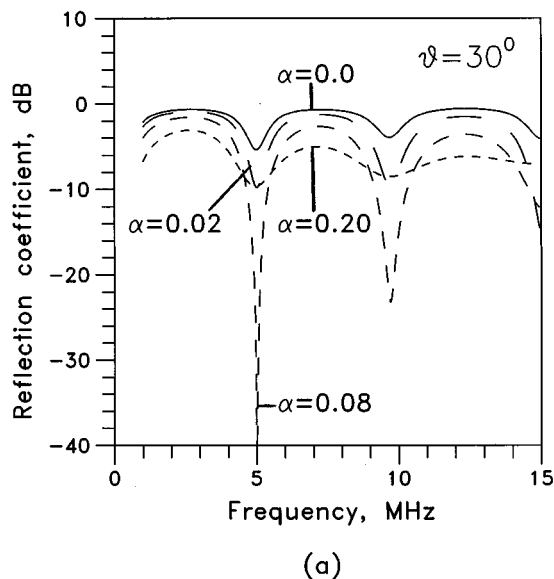


FIG. 15. Amplitude spectra of the ultrasonic wave reflected from the aluminum side of the adhesive layer in the Al/adhesive/steel joint from the aluminum side. Angles of incidence on the Al/adhesive interface are (a) 30° and (b) 60° .

tion in the adhesive can be estimated using a modified relation (11), where instead of R_{12} and R_{32} we substitute reflection coefficients of the SV wave from the Al/adhesive interface at 30° and from the adhesive/steel interface at the SV wave propagation angle in adhesive (9.3°). Estimation for the first spectral minimum yields $\alpha_{\text{critical}} = 0.074$.

When the SV ultrasonic wave impinges on the Al/adhesive interface at 60° both longitudinal and transverse waves are excited inside the adhesive (Fig. 16). The minima observed in the spectrum of the reflected signal [Fig. 15(b)] are due to interference between the SV wave reflected from the Al/adhesive interface and the multiple-reflection signal generated by both SV and longitudinal waves propagating inside the adhesive layer. As a result the minima are not

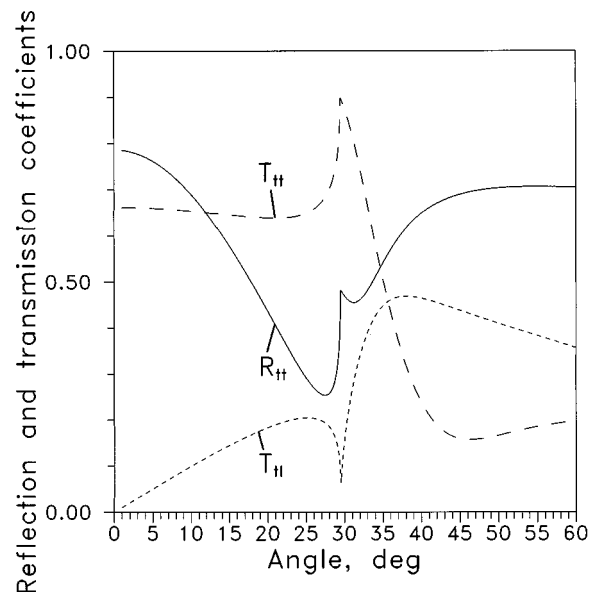


FIG. 16. Reflection and transmission coefficients of a transverse ultrasonic wave at an Al/adhesive interface.

equally spaced and are more difficult to interpret. The effect of anomalous minima deepening with attenuation can still be observed, but the critical attenuation can only be estimated numerically.

B. Experiment

Figure 17 shows amplitude spectra of an ultrasonic wave reflected obliquely from the adhesive layer in the Al/steel adhesive joint. Circles correspond to measurement from the Al side, squares to measurement from the steel side. The incident angle of 16° is greater than the first critical angles in aluminum and steel (13° and 14.4° , respectively), thus only SV waves impinge on the adhesive layer. While the spectral

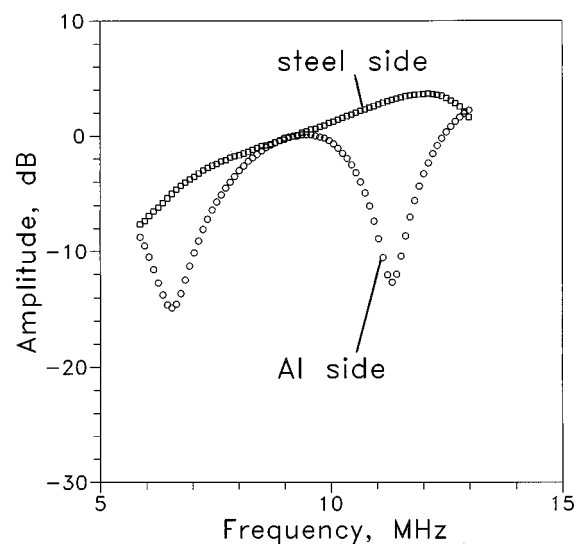


FIG. 17. Experimentally measured amplitude spectra of the ultrasonic signal reflected obliquely from the polymer layer in the Al/steel adhesive joint. The angle of incidence from water is 16° .

minima are well pronounced for insonification from the Al side they disappear from the steel side which is similar to observations at normal incidence (Fig. 10).

VII. CONCLUSIONS

In this paper, the ultrasonic spectroscopy of a layer between two dissimilar materials is studied theoretically and experimentally. It is shown that under certain conditions attenuation increases results in deepening and narrowing of the reflection minimum and that the layer reflectivity measured from the side of the substrate with lower impedance becomes zero at a certain "critical" attenuation. Exact and approximate equations for the critical attenuation and the reflection minimum position are derived. The phenomenon was observed experimentally for both normal incidence and oblique incidence. It is shown that the determination of the attenuation from the measured minima depth is not always unique and thus additional factors such as minimum width of phase spectra should be taken into account. Experiment supports the theoretical predictions. The possibility of using reflected and double transmitted signals for energy loss measurement in the layer is also demonstrated.

- ¹F. H. Chang, P. L. Flynn, D. E. Gordon, and J. R. Bell, "Principles and application of ultrasonic spectroscopy in NDE of adhesive bonds," *IEEE Trans. Sonics Ultrason.* **SU-23**, 334–338 (1976).
- ²P. L. Flynn, "Cohesive bond strength prediction for adhesive joints," *J. Testing Eval.* **7**, 168–171 (1979).
- ³P. A. Meyer and J. L. Rose, "Ultrasonic attenuation effects associated with the physical modeling of adhesive bonds," *J. Appl. Phys.* **48**, 3705–3712 (1976).
- ⁴C. C. H. Guyott and P. Cawley, "The ultrasonic vibration characteristics of adhesive joints," *J. Acoust. Soc. Am.* **83**, 632–640 (1988).
- ⁵C. C. H. Guyott and P. Cawley, "Evaluation of the cohesive properties of adhesive joints using ultrasonic spectroscopy," *NDT Int.* **21**, 233–240 (1988).
- ⁶V. K. Kinra and V. Dayal, "A new technique for ultrasonic nondestructive evaluation of thin specimens," *J. Exp. Mech.* **28**, 288–297 (1988).

- ⁷S. E. Hanneman and V. K. Kinra, "A new technique for ultrasonic non-destructive evaluation of adhesive joints: Part I. Theory," *J. Exp. Mech.* **32**, 323–331 (1992).
- ⁸N. Guo, M. K. Lim, and T. Pialucha, "Measurement of attenuation using a normalized amplitude spectrum," *J. Nondestruct. Eval.* **14**, 9–19 (1995).
- ⁹R. B. Thompson and D. O. Thompson, "Past experience in the development of tests for adhesive bond strength," *J. Adhesion Sci. Technol.* **5**, 583–599 (1991).
- ¹⁰S. I. Rokhlin, M. Hefetz, and M. Rosen, "An ultrasonic interface-wave method for predicting the strength of adhesive bonds," *J. Appl. Phys.* **52**, 2847–2851 (1981).
- ¹¹S. I. Rokhlin, W. Wang, Y. J. Wang, and J. L. Hammill, "Evaluation of Weak Interfaces on Adhesive Joints," in *NDE of Adhesive Bonds and Bondline Topical Proceeding* (ASNT, Columbus, OH, 1989), p. 41.
- ¹²W. Wang and S. I. Rokhlin, "Evaluation of interfacial properties in adhesive joints of aluminum alloys using angle-beam ultrasonic spectroscopy," *J. Adhesion Sci. Technol.* **5**, 647–666 (1991).
- ¹³A. I. Lavrentyev and S. I. Rokhlin, "Models for ultrasonic characterization of environmental degradation of interfaces in adhesive joint," *J. Appl. Phys.* **76**, 4643–4650 (1994).
- ¹⁴L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960), p. 50.
- ¹⁵A. I. Lavrentyev, W. Huang, Y. C. Chu, and S. I. Rokhlin, "Ultrasonic spectroscopy of a layer between dissimilar substrates," in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1995), Vol. 14, pp. 1561–1569.
- ¹⁶W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (Van Nostrand, New York, 1950).
- ¹⁷S. 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).
- ¹⁸S. I. Rokhlin, A. I. Lavrentyev, and B. Li, "Ultrasonic evaluation of environmental durability of adhesive joint," *Res. NDT* **5**, 95–109 (1993).
- ¹⁹S. I. Rokhlin and Y. J. Wang, "Analysis of boundary conditions for elastic wave interaction with an interface between two solids," *J. Acoust. Soc. Am.* **89**, 503–515 (1991).
- ²⁰S. I. Rokhlin and Y. J. Wang, "Equivalent boundary conditions for thin orthotropic layer between two solids: Reflection, refraction, and interface waves," *J. Acoust. Soc. Am.* **91**, 1875–1887 (1992).
- ²¹B. Hosten, "Bulk heterogeneous plane waves propagation through viscoelastic plates and stratified media with large values of frequency domain," *Ultrasonics* **29**, 445–449 (1991).

Bounds and approximations for elastic wave speeds in cubic crystals

Q. H. Zuo and K. D. Hjelmstad

Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 26 August 1996; accepted for publication 24 February 1997)

Unlike an isotropic elastic medium, the wave speeds in an anisotropic medium depend on the direction of wave propagation. In this paper, bounds are developed on the speeds of waves propagating along an arbitrary direction in a cubic crystal through an additive split of the acoustic tensor into two parts, the eigenproperties of which are trivial to determine. The bounds are obtained by applying the minimax property of eigenvalues. Linear estimates of the wave speeds are obtained from a first-order expansion of the eigenvalues of the acoustic tensor, with respect to an anisotropy index, about the value of that index for which anisotropy vanishes. © 1997 Acoustical Society of America. [S0001-4966(97)04306-3]

PACS numbers: 43.35.Cg, 43.20.Bi, 43.20.Jr [CBB]

INTRODUCTION

For a wave propagating along an arbitrary direction in a generally anisotropic material, the wave speeds depend on the direction of propagation and the polarization directions (the directions along which material particles vibrate) can be neither parallel nor perpendicular to the propagation direction. It is generally impossible to find the closed-form expression for the wave speeds and the polarization directions. Pao and Kaul¹ p. 177, called the situation “so complicated that so far no detailed results are available.” Despite the innocent appearance of the acoustic tensor for materials with cubic symmetry (the simplest form of anisotropy), a closed-form expression for the eigenvalues and eigenvectors of the acoustic tensor for an arbitrary propagation direction also is not possible.² Previous work has mostly been devoted to the special propagation directions that admit pure-longitudinal or pure-shear waves (see, for example, Refs. 2–4). For the problem of arbitrary propagation direction, Fedorov⁵ presented a technique for determining equivalent elastic constants of an isotropic material to approximately characterize the acoustic behavior of a cubic crystal. The elastic constants were chosen to minimize the difference between the approximate isotropic and exact anisotropic acoustic tensors over all propagation directions.

For specific values of the material constants and a given wave propagation direction, one can numerically perform the spectral decomposition of the symmetric three by three acoustic tensor. What, then, is the contribution of the approximations presented herein? The compact analytical expressions we present lend crisper and deeper insights into the effects of the material and directional parameters on wave behavior, in particular on the wave speeds. The numerical parameter studies required to gain the same level of insight would be quite tedious to compute and difficult to represent (graphical representation is a difficult task for a problem with five independent parameters). We develop a simple analytical expression for the wave speeds in terms of the anisotropy index and propagation direction. We study the sensitivity of the approximated wave speeds to the change of the propagation direction and show that the directions with extreme sen-

sitivity are those special directions in a crystal along which pure longitudinal and pure shear waves may propagate. Such intriguing features cannot be revealed by numerically computing the eigenvalues and vectors of an acoustic tensor.

I. PRELIMINARIES

For a plane wave propagating in an elastic solid along some arbitrary direction \mathbf{n} , the propagation speed and the polarization direction satisfy the following eigenvalue problem (see, for example, Ref. 6)

$$\mathbf{A}(\mathbf{n}) \cdot \mathbf{u} = \varrho c^2 \mathbf{u}, \quad (1)$$

where \mathbf{u} is the polarization direction, ϱ the density of the material, and c the wave speed. The acoustic tensor, or the Kelvin–Christoffel stiffness,^{2,7} $\mathbf{A}(\mathbf{n})$ in Eq. (1) is defined by

$$\mathbf{A}(\mathbf{n}) = \mathbf{n} \cdot \mathbf{E} \cdot \mathbf{n}, \quad (2)$$

where \mathbf{E} is the fourth-order elasticity tensor of the material, which maps the strain into the stress as

$$\boldsymbol{\sigma} = \mathbf{E} : \mathbf{e}. \quad (3)$$

For an isotropic material the elasticity tensor can be represented as $\mathbf{E} = \bar{\lambda} \mathbf{i} \otimes \mathbf{i} + 2G \mathbf{I}$, where $\bar{\lambda}$ and G are the Lamé constants of the material and \mathbf{i} and \mathbf{I} denote the second- and fourth-order identity tensors, respectively. In this case, the corresponding acoustic tensor has the form $\mathbf{A}(\mathbf{n}) = (\bar{\lambda} + 2G) \mathbf{n} \otimes \mathbf{n} + G(\mathbf{i} - \mathbf{n} \otimes \mathbf{n})$, with distinct eigenvalue $\bar{\lambda} + 2G$, with the corresponding eigenvector \mathbf{n} (longitudinal wave), and repeated eigenvalue G , with corresponding eigenvectors in the plane perpendicular to \mathbf{n} (shear waves). Observe that in an isotropic material the wave speeds are independent of the propagation direction, a well-known fact.

II. CUBIC SYMMETRY

Let λ_1^* , λ_2^* , and λ_3^* represent the three distinct eigenvalues of the fourth-order elasticity tensor defined by^{8,9}

$$\mathbf{E} : \mathbf{N}_i = \lambda_i^* \mathbf{N}_i, \quad (4)$$

where \mathbf{N}_i , for $i = 1, \dots, 6$, are the eigentensors of \mathbf{E} (Ref. 4). Let us define the parameters $\alpha \equiv \frac{1}{3}(\lambda_1^* - \lambda_2^*) + \frac{1}{2}\lambda_3^*$, $\beta \equiv \frac{1}{3}(\lambda_1^* + 2\lambda_2^*) - \frac{1}{2}\lambda_3^*$ and $\gamma \equiv \alpha/\beta$. The acoustic tensor for a cubic crystal can then be expressed as [see, for example, Eq. (24) in Ref. 4]

$$\mathbf{A}(\mathbf{n}) = \frac{1}{2}\lambda_3^* \mathbf{i} + \beta \mathbf{B}. \quad (5)$$

If a Cartesian coordinate system is chosen along the axes of material symmetry (they are defined as the crystal axes), then the components of \mathbf{B} are

$$\begin{bmatrix} n_1^2 & \gamma n_1 n_2 & \gamma n_1 n_3 \\ \gamma n_1 n_2 & n_2^2 & \gamma n_2 n_3 \\ \gamma n_1 n_3 & \gamma n_2 n_3 & n_3^2 \end{bmatrix}. \quad (6)$$

For ease of representing detailed components in constitutive Eq. (3), let us use a fairly standard matrix notation¹⁰

$$\{\sigma\} = [C]\{e\}, \quad (7)$$

where $\{\sigma\}^T = \{\sigma_{11}, \sigma_{22}, \sigma_{33}, \sqrt{2}\sigma_{23}, \sqrt{2}\sigma_{31}, \sqrt{2}\sigma_{12}\}$ and $\{e\}^T = \{e_{11}, e_{22}, e_{33}, \sqrt{2}e_{23}, \sqrt{2}e_{31}, \sqrt{2}e_{12}\}$. The matrix $[C]$ takes the form

$$[C] = \begin{bmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & 2C_{44} \end{bmatrix}, \quad (8)$$

in which the three independent material parameters are C_{11} , C_{12} , and C_{44} . The eigenvalues of the elasticity tensor \mathbf{E} are $\lambda_1^* = C_{11} + 2C_{12}$, $\lambda_2^* = C_{11} - C_{12}$, $\lambda_3^* = 2C_{44}$ (Ref. 10). The constants used in Eq. (5) take the particular form

$$a = C_{12} + C_{44}, \quad \beta = C_{11} - C_{44}, \quad \gamma \equiv \frac{a}{\beta} = \frac{C_{12} + C_{44}}{C_{11} - C_{44}}. \quad (9)$$

The parameter γ measures the degree of anisotropy of the material. For isotropic materials, $2C_{44} = C_{11} - C_{12}$ and, consequently, $\gamma = 1$.

III. BOUNDS ON WAVE SPEEDS

The acoustic tensor $\mathbf{A}(\mathbf{n})$, given in Eq. (5), can be decomposed into

$$\mathbf{A} = \mathbf{A}^I + \mathbf{A}^a, \quad (10)$$

where

$$\begin{aligned} \mathbf{A}^I &\equiv C_{44} \mathbf{i} + (C_{12} + C_{44}) \mathbf{n} \otimes \mathbf{n}, \\ \mathbf{A}^a &\equiv (C_{11} - C_{12} - 2C_{44}) \sum_{i=1}^3 n_i^2 \mathbf{e}_i \otimes \mathbf{e}_i. \end{aligned} \quad (11)$$

The tensor \mathbf{A}^I has the same spectral structure as the acoustic tensor of an isotropic material; it indeed reduces to the latter when cubic symmetry degenerates to isotropy. Let us refer to \mathbf{A}^I as the isotropic counterpart of \mathbf{A} . As such, \mathbf{A}^a can be

viewed as the modification to \mathbf{A}^I due to anisotropy. The tensor \mathbf{A}^a vanishes in the isotropic limit.

The eigenvalues of \mathbf{A} are closely related to those of \mathbf{A}^I and \mathbf{A}^a . Let λ_i , λ_i^I , and μ_i denote the eigenvalues of \mathbf{A} , \mathbf{A}^I , and \mathbf{A}^a , respectively, where all three sets are arranged in nonincreasing order. From the minimax principle of eigenvalue problem,¹¹ we have

$$\mu_3 \leq \lambda_i - \lambda_i^I \leq \mu_1, \quad (12)$$

i.e., the differences between the corresponding eigenvalues of the acoustic tensor \mathbf{A} and the eigenvalues of its isotropic counterpart \mathbf{A}^I are bounded by the smallest eigenvalue of the anisotropic modification \mathbf{A}^a from below and the largest one from above. It follows from Eq. (12) that

$$\lambda_i^I + \mu_3 \leq \lambda_i \leq \lambda_i^I + \mu_1, \quad (13)$$

where λ_i^I and μ_i are readily available due to the simple structures of \mathbf{A}^I and \mathbf{A}^a . In particular,

$$\langle \lambda_i^I \rangle = \langle C_{12} + 2C_{44}, C_{44}, C_{44} \rangle, \quad (14)$$

$$\langle \mu_i \rangle = (C_{11} - C_{12} - 2C_{44}) \langle n_{m1}^2, n_{m2}^2, n_{m3}^2 \rangle, \quad (15)$$

where $m1, m2, m3$ are a permutation of 1, 2, 3 depending on the ordering of the values of n_1, n_2, n_3 and the sign of $C_{11} - C_{12} - 2C_{44}$. Specifically,

$$n_{m1}^2 \geq n_{m2}^2 \geq n_{m3}^2 \quad \text{if } (C_{11} - C_{12} - 2C_{44}) \geq 0, \quad (16)$$

$$n_{m1}^2 \leq n_{m2}^2 \leq n_{m3}^2 \quad \text{if } (C_{11} - C_{12} - 2C_{44}) < 0.$$

Substitution of Eqs. (14) and (15) into Eq. (13) yields

$$\begin{aligned} (C_{12} + 2C_{44}) + (C_{11} - C_{12} - 2C_{44})n_{m3}^2 \\ \leq \lambda_1 \leq (C_{12} + 2C_{44}) + (C_{11} - C_{12} - 2C_{44})n_{m1}^2, \end{aligned} \quad (17)$$

$$\begin{aligned} C_{44} + (C_{11} - C_{12} - 2C_{44})n_{m3}^2 \\ \leq \lambda_3 \leq C_{44} + (C_{11} - C_{12} - 2C_{44})n_{m1}^2. \end{aligned} \quad (18)$$

Recall that the eigenvalues of the acoustic tensor are related to the wave speeds through $\varrho c_i^2 = \lambda_i$. Thus, we have derived upper and lower bounds on the wave speeds in terms of the elastic constants of the cubic crystal and the propagation direction. In the case of isotropy, that is, when $C_{11} - C_{12} = 2C_{44} = 2G$, Eqs. (17) and (18) imply $\lambda_1 = C_{12} + 2C_{44} = \lambda + 2G$ and $\lambda_2 = \lambda_3 = C_{44} = G$, which are just the eigenvalues of the acoustic for an isotropic material. Consider a wave propagating along one of the diagonals of a cubic crystal, i.e., $\langle n \rangle = 1/\sqrt{3} \langle 1, 1, 1 \rangle$. Since $n_{m1}^2 = n_{m2}^2 = n_{m3}^2 = \frac{1}{3}$. Equations (17) and (18) reduce to $\lambda_1 = \frac{1}{3}(C_{11} + 2C_{12} + 4C_{44})$ and $\lambda_2 = \lambda_3 = \frac{1}{3}(C_{11} - C_{12} + C_{44})$, which agrees with established results (e.g., Ref. 6).

Since $0 \leq n_i^2 \leq 1$, looser, but simpler bounds than the ones given in Eqs. (17) and (18) can be established for all directions. For $C_{11} - C_{12} - 2C_{44} \geq 0$ (or $\gamma \geq 1$), we have

$$0 \leq \mu_3 \leq \mu_1 \leq C_{11} - C_{12} - 2C_{44}. \quad (19)$$

A corollary result is that waves in a cubic crystal with $\gamma \leq 1$ cannot be slower than its isotropic counterpart. The substitution of Eqs. (19) and (14) into Eq. (13) yields bounds on the eigenvalues for all directions

$$C_{12} + 2C_{44} \leq \lambda_1 \leq C_{11}, \quad (20)$$

$$C_{44} \leq \lambda_3 \leq \lambda_2 \leq C_{11} - C_{12} - C_{44}.$$

The corresponding bounds on the wave speeds are

$$\begin{aligned} \sqrt{(C_{12} + 2C_{44})/\varrho} &\leq c_1 \leq \sqrt{C_{11}/\varrho}, \\ \sqrt{C_{44}/\varrho} &\leq c_3 \leq c_2 \leq \sqrt{(C_{11} - C_{12} - C_{44})/\varrho}. \end{aligned} \quad (21)$$

Similarly, for $C_{11} - C_{12} - 2C_{44} \leq 0$ (or $\gamma \geq 1$), we have

$$0 \geq \mu_1 \geq \mu_3 \geq C_{11} - C_{12} - 2C_{44}. \quad (22)$$

Consequently, waves in a cubic crystal with $\gamma \geq 1$ cannot be faster than its isotropic counterpart. Furthermore,

$$C_{12} + 2C_{44} \geq \lambda_1 \geq C_{11}, \quad (23)$$

$$C_{44} \geq \lambda_2 \geq \lambda_3 \geq C_{11} - C_{12} - C_{44},$$

and

$$\begin{aligned} \sqrt{(C_{12} + 2C_{44})/\varrho} &\geq c_1 \geq \sqrt{C_{11}/\varrho}, \\ \sqrt{C_{44}/\varrho} &\geq c_2 \geq c_3 \geq \sqrt{(C_{11} - C_{12} - C_{44})/\varrho}. \end{aligned} \quad (24)$$

The bounds given in Eqs. (21) and (24) hold for all propagation directions.

Alternative bounds. Alternative bounds can be established based on splitting the acoustic tensor into the sum of a diagonal and rank-one tensor, respectively, as follows:

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2, \quad (25)$$

where

$$\mathbf{A}_1 = C_{44}\mathbf{i} + (C_{11} - C_{12} - 2C_{44}) \sum_{i=1}^3 n_i^2 \mathbf{e}_i \otimes \mathbf{e}_i, \quad (26)$$

$$\mathbf{A}_2 = (C_{12} + C_{44})\mathbf{n} \otimes \mathbf{n}.$$

The corresponding eigenvalues are $\lambda_i(\mathbf{A}_1) = C_{44} + (C_{11} - C_{12} - 2C_{44})n_i^2$ and $\langle \lambda_i(\mathbf{A}_2) \rangle = (C_{12} + C_{44})\langle 1, 0, 0 \rangle$. It follows from the minimax principal that

$$\begin{aligned} (C_{11} - C_{12} - 2C_{44})n_i^2 + C_{44} \\ \leq \lambda_i \leq (C_{11} - C_{12} - 2C_{44})n_i^2 + (C_{12} + 2C_{44}). \end{aligned} \quad (27)$$

In the case of isotropy $C_{11} = C_{12} + 2C_{44} = \bar{\lambda} + 2G$ and Eq. (27) yields the well-known result

$$G \leq \lambda_i \leq \bar{\lambda} + 2G. \quad (28)$$

The looser bounds implied by Eq. (27) are

$$C_{44} \leq \lambda_i \leq C_{11} \quad \text{for } C_{11} \geq C_{12} + 2C_{44}, \quad (29)$$

$$\begin{aligned} (C_{11} - C_{12} - C_{44}) \leq \lambda_i \leq C_{12} + 2C_{44} \\ \text{for } C_{11} \leq C_{12} + 2C_{44}, \end{aligned} \quad (29)$$

which are compatible with the ones given in Eqs. (20) and (23).

Fedorov⁵ found an acoustic tensor for an isotropic material that provides the best approximation to a cubic crystal over all the propagation directions

$$\begin{aligned} \mathbf{A} \approx (C_{44} + 0.2(C_{11} - C_{12} - 2C_{44}))\mathbf{i} \\ + (C_{12} + C_{44} + 0.4(C_{11} - C_{12} - 2C_{44}))\mathbf{n} \otimes \mathbf{n}. \end{aligned} \quad (30)$$

Since $\mathbf{n} \otimes \mathbf{n}$ is rank-one, the corresponding approximate eigenvalues of \mathbf{A} are

$$\begin{aligned} \langle 0.6C_{11} + 0.4(C_{12} + 2C_{44}), 0.2(C_{11} - C_{12}) + 0.6C_{44}, \\ 0.2(C_{11} - C_{12}) + 0.6C_{44} \rangle, \end{aligned} \quad (31)$$

which obviously fit into the bounds given in either Eq. (20) or (23), depending on the sign of $C_{11} - C_{12} - 2C_{44}$. If Fedorov's isotropic approximation is used in the decomposition of the acoustic tensor in Eq. (10), then the corresponding anisotropic modification becomes

$$\mathbf{A}^a = (C_{11} - C_{12} - 2C_{44}) \left[\sum_{i=1}^3 n_i^2 \mathbf{e}_i \otimes \mathbf{e}_i - 0.2\mathbf{i} - 0.4\mathbf{n} \otimes \mathbf{n} \right].$$

Since the eigenvalues of \mathbf{A}^a cannot be determined easily, one cannot establish the corresponding bounds on the eigenvalues of \mathbf{A} as we have done here.

Example. As an example, consider an aluminum crystal with elastic constants of $C_{11} = 10.82$, $C_{12} = 6.13$, $C_{44} = 2.85$ (Ref. 12). The unit of the constants is 10 GPa. Equations (17) and (18) yield the bounds for a given direction \mathbf{n}

$$11.83 - 1.01n_{m3}^2 \leq \lambda_1 \leq 1.83 - 1.01n_{m1}^2, \quad (32)$$

$$2.85 - 1.01n_{m3}^2 \leq \lambda_3 \leq \lambda_2 \leq 2.85 - 1.01n_{m1}^2,$$

where $n_{m3}^2 = \max\{n_1^2, n_2^2, n_3^2\}$ and $n_{m1}^2 = \min\{n_1^2, n_2^2, n_3^2\}$, since $\gamma = (C_{12} + C_{44})/(C_{11} - C_{44}) = 1.13 > 1$. The corresponding weaker bounds given in Eq. (23) for all the propagation directions are

$$10.82 \leq \lambda_1 \leq 11.83, \quad 1.84 \leq \lambda_3 \leq \lambda_2 \leq 2.85. \quad (33)$$

If the density of aluminum is taken as 2700 kg/m³, then the bounds on the wave speeds in the aluminum crystals are

$$6330 \text{ m/s} \leq c_1 \leq 6619 \text{ m/s}, \quad (34)$$

$$2610 \text{ m/s} \leq c_3 \leq c_2 \leq 3249 \text{ m/s}.$$

The substitution of the material constants into Eq. (31) yields the Fedorov's approximation of the eigenvalues

$$\lambda_1 = 11.22, \quad \lambda_2 = \lambda_3 = 2.64. \quad (35)$$

The corresponding wave speeds are

$$c_1 = 6446 \text{ m/s}, \quad c_2 = c_3 = 3127 \text{ m/s}. \quad (36)$$

Equations (35) and (36) satisfy the bounds given in Eqs. (33) and (34), respectively.

IV. APPROXIMATIONS OF WAVE SPEEDS

It follows from Eq. (5) that \mathbf{A} and \mathbf{B} share the same eigenvectors and their eigenvalues are related by

$$\lambda(\mathbf{A}) = C_{44} + (C_{11} - C_{44})\lambda(\mathbf{B}), \quad (37)$$

where we have noted that $\lambda_3^* = 2C_{44}$ and $\beta = C_{11} - C_{44}$. It suffices to study the eigenvalues of \mathbf{B} to find the wave speeds in a cubic crystal. Since \mathbf{B} is a tensor that depends analytically on the anisotropy index γ , defined in Eq. (9), its eigenvalues are also analytic functions of γ . For values of γ close to one, i.e., the material is weakly anisotropic (for example, $\gamma = 0.70$ and 1.13 for titanium carbide and aluminum crystals, respectively), then, by Rellich's theorem on Taylor ex-

pansions of eigenvalues of perturbed linear operators¹³ the first-order approximation of the eigenvalues of \mathbf{B} , in terms of the γ , is

$$\lambda_i(\gamma) \approx \lambda_i(1) + (\gamma - 1)\lambda'_i(1), \quad (38)$$

where $\lambda_i(1)$ and $\lambda'_i(1)$ are the eigenvalues and their derivatives, respectively, of \mathbf{B} evaluated at $\gamma = 1$. It follows from Eq. (6) that at $\gamma = 1$,

$$\mathbf{B} = \mathbf{n} \otimes \mathbf{n} \quad (39)$$

with eigenvalues $\langle \lambda_i(1) \rangle = \langle 1, 0, 0 \rangle$. The eigenvectors are \mathbf{n} , and any vector that lies in the plane normal to \mathbf{n} . The derivative of the distinct eigenvalue is¹⁴

$$\lambda'_1 = \mathbf{n} \cdot \mathbf{B}' \cdot \mathbf{n}, \quad (40)$$

since \mathbf{n} is the eigenvector of \mathbf{B} corresponding to the eigenvalue $\lambda_1 = 1$. The tensor \mathbf{B}' is the derivative of \mathbf{B} with respect to the parameter γ , which can be found by differentiating Eq. (6)

$$\mathbf{B}' = \mathbf{n} \otimes \mathbf{n} - \sum_{i=1}^3 n_i^2 \mathbf{e}_i \otimes \mathbf{e}_i. \quad (41)$$

Substituting the above equation into Eq. (40) yields

$$\lambda'_1 = 1 - \sum_{i=1}^3 n_i^4. \quad (42)$$

Therefore the linear approximation of the first eigenvalue of \mathbf{B} is

$$\lambda_1(\gamma) \approx \gamma + (1 - \gamma) \sum_{i=1}^3 n_i^4. \quad (43)$$

For the repeated eigenvalues, Eq. (40) is not valid since the eigenvectors are not unique. Let $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ be orthogonal unit vectors spanning the plane perpendicular to \mathbf{n} . Following Mills-Curran,¹⁵ the derivatives λ'_2 and λ'_3 are the eigenvalues of $\boldsymbol{\Phi}^T \cdot \mathbf{B}' \cdot \boldsymbol{\Phi}$, where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2]$ is the subspace spanned by the eigenvectors corresponding to the repeated zero eigenvalue. Without loss of generality, assume $n_1 \neq 1$. Then one particular choice of the eigenvectors corresponding to the repeated zero eigenvalues is

$$\boldsymbol{\phi}_1 = \frac{1}{\sqrt{1 - n_1^2}} (-n_3 \mathbf{e}_2 + n_2 \mathbf{e}_3), \quad (44)$$

$$\boldsymbol{\phi}_2 = \frac{1}{\sqrt{1 - n_1^2}} ((1 - n_1^2) \mathbf{e}_1 - n_1 n_2 \mathbf{e}_2 - n_1 n_3 \mathbf{e}_3).$$

Note that $\boldsymbol{\phi}_i \cdot \mathbf{n} = 0$ and $\boldsymbol{\phi}_i \cdot \boldsymbol{\phi}_j = \delta_{ij}$. The components of $\boldsymbol{\Phi}^T \cdot \mathbf{B}' \cdot \boldsymbol{\Phi}$ are found to be

$$-\frac{1}{1 - n_1^2} \begin{bmatrix} 2n_2^2 n_3^2 & (n_2^2 - n_3^2) \nu(\mathbf{n}) \\ (n_2^2 - n_3^2) \nu(\mathbf{n}) & 2n_1^2 (1 - n_1^2 - \tau(\mathbf{n})) \end{bmatrix}, \quad (45)$$

where $\tau(\mathbf{n}) \equiv n_1^2 n_2^2 + n_2^2 n_3^2 + n_3^2 n_1^2$ and $\nu(\mathbf{n}) \equiv n_1 n_2 n_3$. The derivatives of the repeated zero eigenvalues of \mathbf{B} are the eigenvalues of $\boldsymbol{\Phi}^T \cdot \mathbf{B}' \cdot \boldsymbol{\Phi}$

$$\lambda'_2, \lambda'_3 = -\tau(\mathbf{n}) \pm \sqrt{\tau^2(\mathbf{n}) - 3\nu^2(\mathbf{n})}. \quad (46)$$

The above equation shows that the derivatives of the repeated zero eigenvalues of \mathbf{B} are unique for a given direction \mathbf{n} , even though the corresponding eigenvectors are not unique. Note that although preference was given to n_1 when we constructed $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, the derivatives of the corresponding repeated eigenvalues are completely symmetric in n_1 , n_2 , and n_3 , as expected. Substituting Eq. (46) into Eq. (38) yields the linear approximations of λ_2 and λ_3 as

$$\lambda_2(\gamma), \lambda_3(\gamma) \approx (1 - \gamma)(\tau(\mathbf{n}) \mp \sqrt{\tau^2(\mathbf{n}) - 3\nu^2(\mathbf{n})}). \quad (47)$$

Equations (43) and (47) give the closed-form approximations of the eigenvalues of \mathbf{B} for arbitrary propagation direction \mathbf{n} when the material is weakly anisotropic. Substituting Eqs. (43) and (47) into Eq. (37) gives the approximations of the eigenvalues of \mathbf{A} , which are proportional to the squares of the wave speeds.

The sum of the eigenvalues of a tensor equals the trace of the tensor. It follows from Eq. (6) that

$$\text{trace}(\mathbf{B}) = \sum_{i=1}^3 n_i^2 = 1. \quad (48)$$

Therefore, the sum of the three eigenvalues of tensor \mathbf{B} is 1

$$\sum_{i=1}^3 \lambda_i(\gamma) = 1. \quad (49)$$

It is straightforward to show that our first-order approximations satisfy Eq. (49). To wit, the sum of the right-hand sides of Eqs. (43) and (47) is

$$\gamma + (1 - \gamma) \left(\sum_{i=1}^3 n_i^4 + 2\tau(\mathbf{n}) \right) = 1 \quad (50)$$

since the term in large parentheses is $(n_1^2 + n_2^2 + n_3^2)^2$.

The accuracy of the linear approximations of eigenvalues of \mathbf{A} depends on the degree of anisotropy of the crystal, but the propagation direction also has an important effect on the accuracy. In a cubic crystal there are three distinct directions along which pure-longitudinal waves can propagate: the crystal axes, the face diagonals, and the body diagonals.³ Let us first consider a wave along a crystal axis, say, $\mathbf{n} = \mathbf{e}_1$, or $\langle n \rangle = \langle 1, 0, 0 \rangle$. Then approximate expressions (43) and (47) yield $\langle \lambda(\gamma) \rangle \approx \langle 1, 0, 0 \rangle$. Similarly, $\langle n \rangle = (1/\sqrt{2}) \times \langle 0, 1, 1 \rangle$ and $\langle n \rangle = (1/\sqrt{3}) \langle 1, 1, 1 \rangle$ yields $\langle \lambda(\gamma) \rangle \approx \frac{1}{2} \langle 1 + \gamma, 0, 1 - \gamma \rangle$ and $\langle \lambda(\gamma) \rangle \approx \frac{1}{3} \langle 1 + 2\gamma, 1 - \gamma, 1 - \gamma \rangle$, respectively. Tensor \mathbf{B} defined in Eq. (6) is particularly simple for those pure wave directions and it can be easily verified that the first-order approximations given here are actually exact (see, for example, Ref. 4). In addition to those three “privileged” directions, pure-shear waves can propagate along the crystal faces (i.e., $n_1 n_2 n_3 = 0$), and directions which form equal angles with two crystal axes [i.e., $(n_1^2 - n_2^2)(n_2^2 - n_3^2) \times (n_3^2 - n_1^2) = 0$]. Suppose that $n_1 = 0$, then $n_2^2 + n_3^2 = 1$ and Eqs. (43) and (47) reduce to $\langle \lambda(\gamma) \rangle \approx \langle 1 + 2(\gamma - 1)n_2^2 n_3^2, 0, 2(1 - \gamma)n_2^2 n_3^2 \rangle$. The exact solution is $\langle \lambda(\gamma) \rangle = \frac{1}{2} \langle 1 + \sqrt{1 + 4(\gamma^2 - 1)n_2^2 n_3^2}, 0, 1 - \sqrt{1 + 4(\gamma^2 - 1)n_2^2 n_3^2} \rangle$ (Ref. 4). For weakly anisotropic crystals (γ close to 1), $\sqrt{1 + 4(\gamma^2 - 1)n_2^2 n_3^2} \approx 1 + 2(\gamma^2 - 1)n_2^2 n_3^2 \approx 1 + 4(\gamma - 1)n_2^2 n_3^2$ and the exact solution reduces to our first-order approxima-

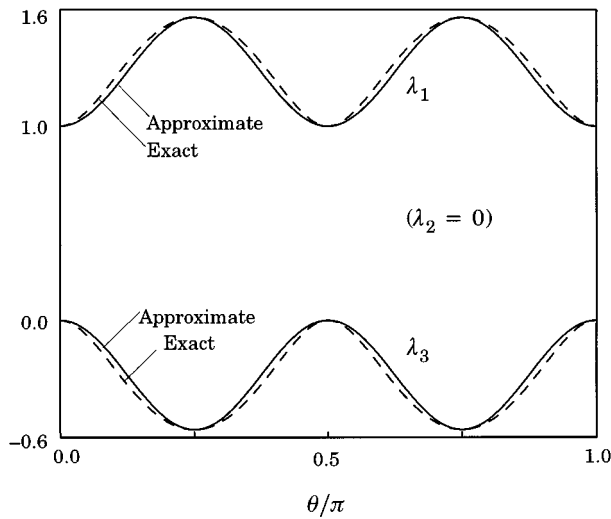


FIG. 1. Comparison of eigenvalues as functions of wave propagation direction.

tion. It is also interesting to note that when $n_2^2 n_3^2 = 0$ or $n_2^2 n_3^2 = \frac{1}{4}$, the first-order approximation is also exact since pure-longitudinal waves can propagate along those directions.

Let θ denote the angle between the propagation direction and the crystal axis \mathbf{e}_2 . Then $n_2^2 n_3^2 = \cos^2 \theta \sin^2 \theta$, and for a given material the eigenvalues of \mathbf{B} are functions of the angle θ only. Figure 1 gives a comparison of such functions for the linear approximations and the exact solutions for a copper crystal with anisotropy index $\gamma = 2.12$. Of the commonly used metal crystals that have cubic symmetry, copper is the most anisotropic.¹² It can be seen that the first-order linear approximations compare well with the exact solutions. The differences between the linear approximations and the exact solutions are zero at $\theta = 0$, increase gradually with the angle θ , reach a maximum at $\theta = 0.35$ rad, then monotonically decrease to zero at $\theta = \pi/4$. The curves are symmetric about $\theta = \pi/4$ because of the cubic symmetry of the material.

V. EXTREMIZING DIRECTIONS

It follows from Eqs. (42) and (46) that the derivatives of the eigenvalues of \mathbf{A} depend on the direction \mathbf{n} . The directions along which the derivative of the first eigenvalue reaches an extremum can be found by extremizing the Lagrangian

$$\mathcal{L}_1(\mathbf{n}, L) \equiv 1 - \sum_{i=1}^3 n_i^4 + 2L(\mathbf{n} \cdot \mathbf{n} - 1), \quad (51)$$

where L is the Lagrange multiplier. Note that if $\|\mathbf{n}\| = 1$ then $\mathcal{L}_1 = \lambda_1'$. Setting the derivative of \mathcal{L}_1 with respect to n_j equal to zero yields

$$-4n_j^3 + 4Ln_j = 0. \quad (52)$$

The solution is $n_j = 0$, or $L = n_j^2$ for each j , which yields three distinct directions in a cubic crystal: $\langle \mathbf{n} \rangle = \langle 1, 0, 0 \rangle$; $(1/\sqrt{2})\langle 0, 1, 1 \rangle$ and $(1/\sqrt{3})\langle 1, 1, 1 \rangle$, the directions along which pure-longitudinal waves may propagate. The corresponding

values of λ_1' along those directions are 0 , $\frac{1}{2}$, and $\frac{2}{3}$, respectively. Therefore, the body diagonals yield the largest value of λ_1' , and the crystal axes yield the smallest value.

Similarly, the directions along which the derivative of the second eigenvalue reaches an extremum can be found by extremizing the Lagrangian

$$\mathcal{L}_2(\mathbf{n}, M) \equiv -\tau(\mathbf{n}) + \sqrt{\tau^2(\mathbf{n}) - 3\nu^2(\mathbf{n})} + M(\mathbf{n} \cdot \mathbf{n} - 1). \quad (53)$$

The solutions are found to be those that extremize \mathcal{L}_1 , as well as directions lying in one of the crystal faces, i.e., characterized by $n_1 n_2 n_3 = 0$. Two distinct values of λ_2' are realized at those directions: 0 along the crystal faces and $-\frac{1}{3}$ along the body diagonals. It is interesting to note that the directions for which λ_2' reaches extrema coincide with those for pure-shear waves.⁴ The directions that extremize λ_3' are found to be those that extremize λ_1' , and $\langle \mathbf{n} \rangle = \frac{1}{2}\langle \sqrt{2}, 1, 1 \rangle$ which also admits a pure-shear wave. The values of λ_3' at those directions are $-\frac{1}{3}$ for $\langle \mathbf{n} \rangle = (1/\sqrt{3})\langle 1, 1, 1 \rangle$; $-\frac{3}{8}$ for $\frac{1}{2}\langle \sqrt{2}, 1, 1 \rangle$, and $-\frac{1}{2}$ for $(1/\sqrt{2})\langle 0, 1, 1 \rangle$, respectively. Therefore, λ_3' reaches its maximum along the body diagonals, and the minimum along the face diagonals.

VI. CONCLUSIONS

Speeds of an elastic wave propagating in cubic crystals along arbitrary directions have been studied in this paper. Two sets of bounds on the speeds have been established by using the minimax principle of eigenvalues. The first bounds are based on a split of the acoustic tensor into isotropic and anisotropic parts. The second bounds are based on decomposing the tensor into a full-rank tensor and a rank-one modification. When cubic symmetry degenerates to isotropy, the bounding inequality in the first set becomes equality and yields the longitudinal and shear wave speeds of the isotropic material. Also, for waves propagating along a body diagonal the upper and lower bounds of the first set coalesce at the exact values of the wave speeds.

For weakly anisotropic crystals, linear approximations for the speeds have been developed by expanding the eigenvalues of the acoustic tensor around the isotropic state in terms of the anisotropy index. The derivatives of the eigenvalues at the isotropic state can be easily computed because the eigenvalues and eigenvectors are either known or can be conveniently constructed and the derivative of the acoustic tensor is readily available. Along the directions that propagate pure-longitudinal waves, such as the body diagonal, the face diagonals, and the crystal axes, our first-order approximations give the exact wave speeds. For the directions lying in one of the crystal faces (in which pure-shear waves may propagate), the approximations differ from the exact solution. Comparisons have been made for copper, which is the most anisotropic among commonly used metals, for all propagation directions showing that the first-order approximations agree quite well with the exact solutions.

ACKNOWLEDGMENTS

This paper was prepared from a study conducted in the Center of Excellence for Airport Pavement Research which

is funded in part by the Federal Aviation Administration under Research Grant No. 95-C-001. The opinions expressed in this paper are those of the authors and do not necessarily reflect the official views and policies of the FAA.

- ¹Y.-H. Pao and R. K. Kaul, "Waves and Vibrations in Isotropic and Anisotropic Plates," in *R. D. Mindlin and Applied Mechanics*, edited by G. Herrmann (Pergamon, New York, 1974), pp. 149–195.
- ²M. J. P. Musgrave, *Crystal Acoustics* (Holden-Day, San Francisco, CA, 1970).
- ³R. N. Thurston, "Waves in Solids," in *Encyclopedia of Physics*, edited by C. Truesdell (Springer-Verlag, New York, 1974), Vol. VIa/4, pp. 109–308.
- ⁴Q. H. Zuo and H. L. Schreyer, "A note on pure-longitudinal and pure-shear waves in cubic crystals," *J. Acoust. Soc. Am.* **98**, 580–583 (1995).
- ⁵F. I. Fedorov, *Theory of Elastic Waves in Crystals* (Plenum, New York, 1968).
- ⁶Ph. Boulanger and M. Hayes, "Waves in Elastic Media," in *Stability and Wave Propagation in Fluids and Solids*, edited by G. P. Galdi (Springer-Verlag, New York, 1995), pp. 1–38.
- ⁷B. A. Auld, *Acoustic Fields and Waves in Solids*, (Krieger, Melbourne, FL, 1990), 2nd ed., Vol. 1, Chap. 7, pp. 191–264.
- ⁸S. Sutcliffe, "Spectral Decomposition of the Elasticity Tensor," *J. Appl. Mech.* **59**, 762–773 (1992).
- ⁹H. L. Schreyer and Q. H. Zuo, "Anisotropic Yield Surfaces Based on Elastic Projection Operators," *J. Appl. Mech.* **62**, 780–785 (1995).
- ¹⁰M. M. Mehrabadi, S. C. Cowin, and C. O. Horgan, "Strain Energy Density Bounds for Linear Anisotropic Materials," *J. Elast.* **30**, 191–196 (1993).
- ¹¹J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).
- ¹²R. W. Hertzberg, *Deformation and Fracture Mechanics of Engineering Materials* (Wiley, New York, 1987), 3rd ed., Chap. 1, pp. 3–47.
- ¹³T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1984).
- ¹⁴R. L. Fox and M. P. Kapoor, "Rates of Change of Eigenvalues and Eigenvectors," *AIAA J.* **6**, 2426–2429 (1968).
- ¹⁵W. C. Mills-Curran, "Calculation of Eigenvector Derivatives for Structures with Repeated Eigenvalues," *AIAA J.* **26**, 867–871 (1988).

Visualization of acoustic particle interaction and agglomeration: Theory evaluation

Thomas L. Hoffmann^{a)} and Gary H. Koopmann

Center for Acoustics and Vibration, The Pennsylvania State University, 157 Hammond Building,
University Park, Pennsylvania 16802

(Received 26 July 1995; revised 17 December 1996; accepted 7 January 1997)

In this paper experimentally observed trajectories of particles undergoing acoustically induced interaction and agglomeration processes are compared to and validated with numerically generated trajectories based on existing agglomeration theories. Models for orthokinetic, scattering, mutual radiation pressure, and hydrodynamic particle interaction are considered in the analysis. The characteristic features of the classical orthokinetic agglomeration hypothesis, such as collision processes and agglomerations due to the relative entrainment motion, are not observed in the digital images. The measured entrainment rates of the particles are found to be consistently lower than the theoretically predicted values. Some of the experiments reveal certain characteristics which may possibly be related to mutual scattering interaction. The study's most significant discovery is the so-called tuning fork agglomeration [T. L. Hoffmann and G. H. Koopmann, *J. Acoust. Soc. Am.* **99**, 2130–2141 (1996)]. It is shown that this phenomenon contradicts the theories for mutual scattering interaction and mutual radiation pressure interaction, but agrees with the acoustic wake effect model in its intrinsic feature of attraction between particles aligned along the acoustic axis. A model by Dianov *et al.* [*Sov. Phys. Acoust.* **13** (3), 314–319 (1968)] is used to describe this effect based on asymmetric flow fields around particles under Oseen flow conditions. It is concluded that this model is consistent with the general characteristics of the tuning fork agglomerations, but lacks certain refinements with respect to accurate quantification of the effect. © 1997 Acoustical Society of America. [S0001-4966(97)04205-7]

PACS numbers: 43.35.Ty [HEB]

INTRODUCTION

High intensity sound waves propagating through an aerosol induce a variety of interaction effects between the suspended particles, some of which yield collisions and agglomerations. These agglomerations, in turn, move the aerosol's particle size distribution toward larger sizes, an effect that may be utilized for certain industrial applications with a need for small particle removal (e.g., cyclone filters, electrostatic precipitators, etc.). For a detailed introduction to this topic and a comprehensive literature review, the reader is referred to our earlier paper.¹ The first part of that work included a discussion of existing acoustic agglomeration theories as well as an experimental visualization study which provided new insights into particle interaction and agglomeration processes. More specifically, a numerical evaluation was carried out on four different particle interaction theories: orthokinetic particle entrainment, scattering interaction, and the hydrodynamic effects related to mutual radiation pressure, and acoustic wake effect. This evaluation produced several important results. Orthokinetic particle entrainment together with scattering interaction led to zones of repulsion or attraction for particles aligned, respectively, along the acoustic axis or perpendicular to it (on-axis or off-axis orientation). The attraction zone (off-axis) was found to be energetically unstable, resulting in particle movement towards the

repulsive areas (on-axis). Of the two hydrodynamic effects under consideration, only the acoustic wake effect exposed on-axis particle attraction and extended its range of interaction to particle separation distances of interest ($\sim 200\text{--}300\text{ }\mu\text{m}$). The theory for mutual radiation pressure interaction predicted on-axis particle repulsion, in disagreement with some of the experimental results. In addition, the effect was shown to be significant only in the immediate vicinity of the particles. There, however, the two hydrodynamic effects had opposing attraction–repulsion patterns, leading to the hypothesis of a captive force field which could keep the particles in a stable configuration at a certain separation distance.

The second part of our earlier paper presented new experimental observations of microscopic particle trajectories displaying interaction and agglomeration phenomena in a high intensity acoustic field. The recorded digitized images revealed a number of different interaction phenomena as well as one distinct attraction pattern, the so-called tuning fork agglomeration. In the range of parameters investigated, the latter phenomenon appeared to be the principal agglomeration mechanism between particles of equal or almost equal size. In the present paper, the characteristic features of these experimental findings are evaluated, analyzed, and contrasted with theoretical results. The objective is to gain a qualitative and quantitative understanding of the observed interaction mechanisms which leads to a critical evaluation of the different agglomeration models.

^{a)}Present address: Instituto de Acústica (CSIC), Serrano 144, 28006 Madrid, Spain.

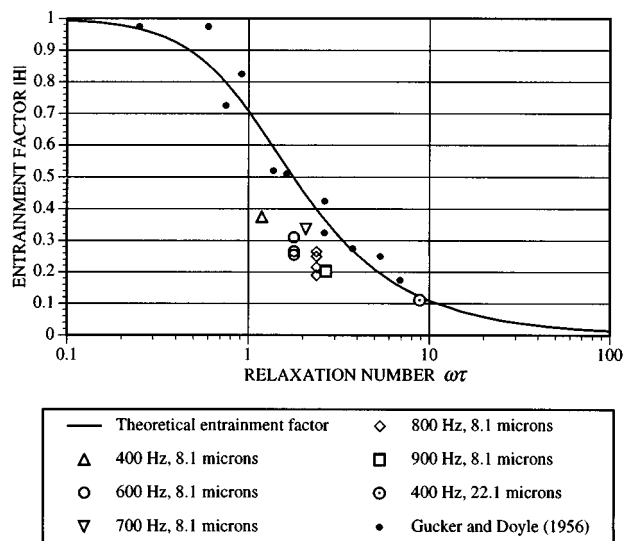


FIG. 1. Theoretical entrainment factor as a function of the relaxation number $\omega\tau$ in comparison to the experimental results of Gucker and Doyle and those of the present paper.

I. COMPARISON BETWEEN THEORY AND EXPERIMENT

A. Acoustic particle entrainment

Acoustic entrainment by viscous forces is the most fundamental interaction phenomenon occurring between an acoustic wave and a suspended particle. The importance of this effect is that it provides a basis for the classical or orthokinetic agglomeration theory discussed in the following section. The origin of the acoustic entrainment theory dates back to König³ (1891). Since then, the oscillatory motion of particles in an acoustic field has been studied by various authors. In 1959, Gucker and Doyle⁴ determined entrainment factors of droplets experimentally as a function of their radii (droplets in the range of 0.8–3.9 μm). Later, Temkin⁵ represented Gucker and Doyle's data as a function of the nondimensional relaxation number $\omega\tau$, with τ being the relaxation time $\frac{2}{3}a^2\rho_p/\mu_0$, ω the circular frequency, μ_0 the dynamic viscosity of the gas, a the particle radius, and ρ_p the particle density. The conversion of the data into a nondimensional form enables common representation of entrainment factors from experiments with different particle sizes, particle densities, and gas viscosities. In Fig. 1, Gucker and Doyle's nondimensionalized results are compared with the theoretical entrainment factor derived from the Basset–Boussinesq–Oseen (BBO) equation^{1,6} and the experimental numbers obtained in the present investigation. Our experimental results were determined from the observed displacement amplitudes at five different frequencies (400, 600, 700, 800, and 900 Hz) for two different particle sizes (8.1 and 22.1 μm) and at varying acoustic velocity amplitudes (0.64–1.70 m/s). As a general trend, the values of the current work are found to be shifted toward lower entrainment rates. This shift is more pronounced at smaller relaxation numbers. On the other hand, Gucker and Doyle's data agree quantitatively with the theoretical curve. However, as will be pointed out below, the experimental approach of the two studies differs signifi-

cantly. While Gucker and Doyle limit their experiments to a single frequency at one acoustic velocity level, the data of this study are compiled from different results obtained in the abovementioned parameter range. Moreover, the particle size in the current study is a well-defined quantity because of the employment of monodisperse glass microsphere size standards with certified diameters. By contrast, Gucker and Doyle determine particle sizes from statistical evaluation of rate-of-fall experiments. Moreover, their droplet size measurements were not carried out simultaneously with the vibration amplitude determination. Therefore, even though Gucker and Doyle's data may seem to give a better fit to the theoretical curve, their results should be evaluated carefully in the light of these experimental limitations. On the other hand, the values obtained in this study display consistency and small standard deviations in the covered range of frequencies, amplitudes, and particle sizes. Although shifted toward lower entrainment values, the general behavior of the theoretical function is well reflected by the data. Thus, the new experimental data confirm the trend described by the BBO equation, but disagree in the exact quantitative evaluation of the entrainment rates. For verification of these new results, a forthcoming paper⁷ will focus on more detailed measurements with a different experimental setup, covering a wider range of parameters.

B. Orthokinetic particle interaction

The particle entrainment function of Fig. 1 is directly related to the orthokinetic acoustic agglomeration hypothesis. This classical agglomeration model implies that particles of different sizes are entrained into the oscillatory flow of the acoustic field at different rates of motion. The more stationary, larger particles function as collection nuclei with which the smaller particles collide and agglomerate due to their stronger entrainment in the fluid motion. Therefore, orthokinetic agglomeration is, by definition, limited to pairs of particles of different sizes. The visualization study presented in our earlier paper¹ revealed digitized images of typical "orthokinetic" entrainment situations, showing large stationary particles and small strongly oscillating ones. However, due to amplitude limitations of the loudspeakers employed, only the experiments at the lowest frequencies (400 Hz) generated particle displacements within or exceeding the range of the particle separation distances. Displacements of this order are a necessary condition for collision processes as predicted by the orthokinetic hypothesis. At higher frequencies (600–1000 Hz) the maximum observed particle displacements dropped far below the inter-particle separation; thus, no conclusions could be drawn from these experiments with respect to orthokinetic agglomeration. In this context it should be noted, however, that at constant acoustic velocity the displacement level decays with 6 dB/frequency doubling. This implies that with increasing frequency orthokinetic agglomeration requires either smaller particle separation (i.e., larger number density of the aerosol) or higher velocity amplitudes (which for technical reasons are generally not achievable). Therefore, orthokinetic particle interaction must be more pronounced at lower frequencies, the range in which the present experiments were conducted.

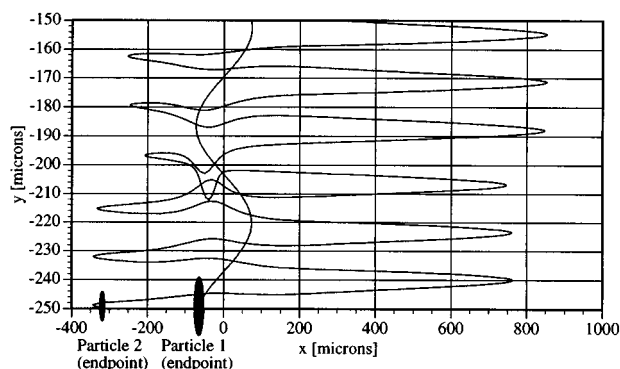


FIG. 2. Numerically determined trajectories of two spherical particles (diameters 10 and 20 μm) under the influence of orthokinetic, scattering and gravitational effects (the acoustic sound field vector is aligned horizontally), frequency 400 Hz, acoustic velocity amplitude 5 m/s, particle density 2600 kg/m^3 (quartz).

For the analysis, a large number of digital images were taken showing polydisperse size distributions of quartz particles ($<50 \mu\text{m}$), as well as bidisperse particle mixtures consisting of 8- and 21- μm glass microspheres. The particles were excited in a homogeneous acoustic field with a velocity of about 1.2 m/s rms at 400 Hz ($L_v \approx 148 \text{ dB}$). Even though a considerable number of images were available for the evaluation,¹ not one single incidence of collision or agglomeration could be observed. This result indicates that at least under the experimental conditions of this study orthokinetic agglomeration is a rare phenomenon. However, no attempt should be made to generalize these results because of the mentioned experimental limitations on the investigated parameter range. Further experiments are necessary to elucidate orthokinetic particle interaction at higher frequencies and with larger number densities.

C. Mutual scattering interaction

If we consider two close-by particles we find that they are not only entrained into the primary acoustic field, but also oscillate with the fluid motion of the scattered wave fields. This effect, termed mutual scattering interaction, was proposed by Song^{8,9} as an important contributor for effective refill of the emptied agglomeration volumes. In our earlier paper¹ we investigated this interaction phenomenon by analyzing numerically generated particle trajectories. It became clear that the effect fails to generate significant attractive motion between particles which are oriented with their common axis perpendicular to the acoustic field vector (off-axis orientation). The particles were either repelled in the on-axis direction or, if positioned off-axis, drawn into the repulsive zones. If relative gravitational motion was taken into account, the time of interaction was reduced to the short period of closest particle approach. Figure 2 shows an example of a numerically generated trajectory displaying the path of a pair of particles (diameters 10 and 20 μm) based on orthokinetic, scattering, and gravitational effects. The appearance of the larger particle leads only to minor deviations of the small particle trajectory during the short timespan of about five acoustical cycles. Approach patterns as suggested in Song's

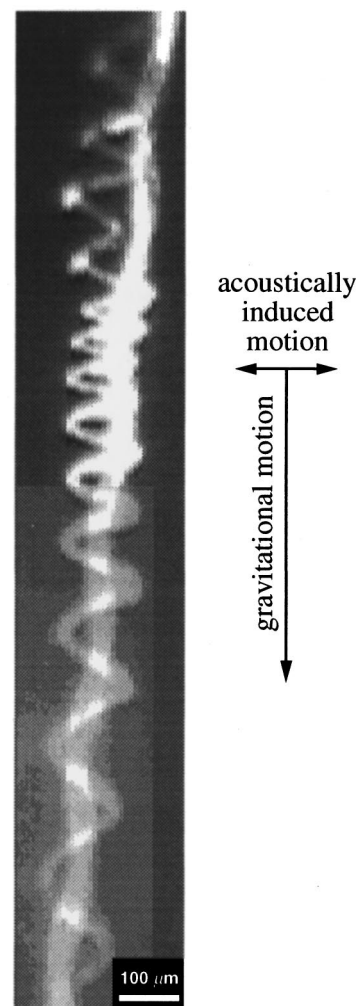


FIG. 3. Interaction of two quartz particles (diameters $<50 \mu\text{m}$) dropping vertically; frequency 400 Hz, acoustic velocity amplitude 1.70 m/s, particle density $\approx 2600 \text{ kg/m}^3$ (quartz).

model^{1,8,9} are not confirmed if gravity effects are included in the computations. In some of the visualization experiments, interaction phenomena were observed that possibly relate to the trajectories of Fig. 2. Especially in the experiments at 400 Hz, where single oscillations of the particles were clearly visible, the sinusoidal shape of the trajectories changed under the influence of a neighboring larger particle. Generally, this caused the vertical particle velocity to diminish during a short period of time. After the interaction the particle returned to its initial oscillatory motion. Figure 3 shows an example of this effect. Here a larger particle “falls” from above into the path of a smaller, more entrained particle. During the time of closest approach the motion of the smaller particle in the vertical direction slows down and increases again after the larger particle has passed by. While these results exhibit certain features which can possibly be related to mutual scattering effects, no direct evidence is found to support Song's hypothesis of scattering interaction as a sufficient refill mechanism for the acoustic agglomeration volume.

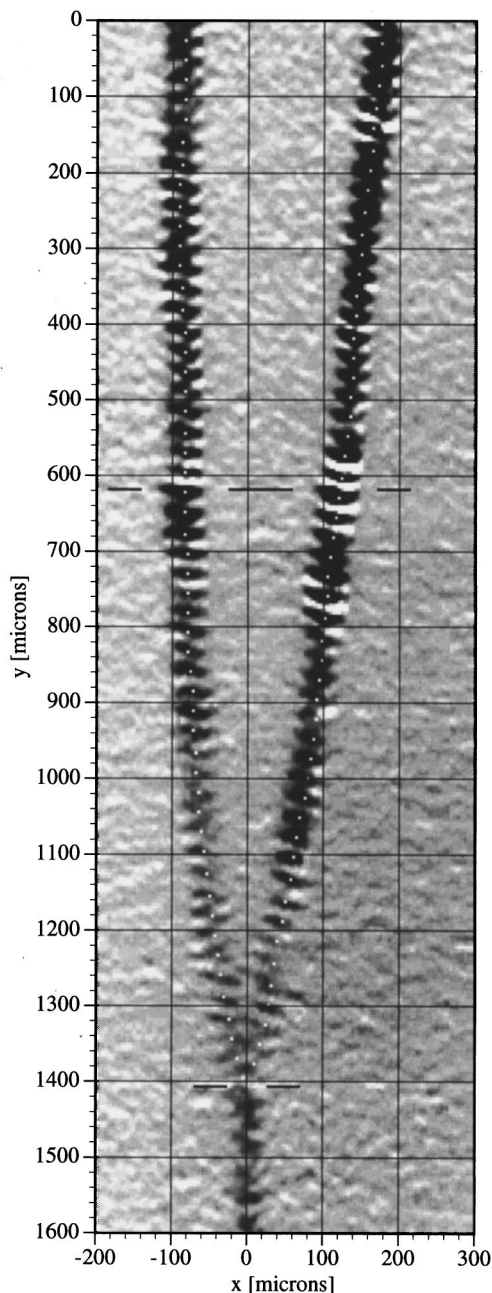


FIG. 4. Time-resolved reconstruction of a tuning fork agglomeration in periods of 1.1 ms with two $8.1\text{-}\mu\text{m}$ microspheres dropping vertically (particle density 2400 kg/m^3 —soda lime glass); the acoustic sound field vector is aligned horizontally; acoustic velocity amplitude 0.75 m/s ; frequency 900 Hz ; vertical velocity 23.5 mm/s .

D. Comparison of the observed tuning fork agglomerations and the acoustic wake effect¹⁰

In the experimental study¹ trajectories of acoustically excited glass microspheres were visualized by means of a CCD video camera connected to an image processing and analysis system. The acquired images revealed different types of acoustically induced particle interaction and attraction patterns. An important result of this visualization was the discovery of tuning fork shaped attraction patterns at higher frequencies and amplitudes (e.g., see Fig. 4). These repeatedly observed trajectories are the result of two differ-

ent forces acting on the particles, the gravitational force with vertical orientation and an acoustically induced hydrodynamic force which moves the particles horizontally in the direction of the acoustic field vector. Experiments with $8.1\text{-}\mu\text{m}$ microspheres showed that the number of such interactions increased drastically at frequencies higher than 600 Hz and at acoustic velocities of 0.45 m/s and above. It was pointed out that the tuning fork patterns, which are related to the acoustic wake effect, evidence an important agglomeration mechanism, especially for the interaction of equally or almost equally sized particles.

For the analysis of the experimental data and the comparison between theory and experiment it is necessary to emphasize certain implicit features of the captured video images. Figure 4 shows a typical tuning fork attraction pattern obtained in an experiment at 900 Hz with an acoustic velocity amplitude of 0.75 m/s . The trajectories are reconstructed from three subsequently captured video frames (each of which represents a timespan of $1/30\text{ s}$). Two horizontal black cutlines indicate the endpoints of the trajectories in one frame or, equally, the starting points of the trajectories in the next one. This implies that it is possible to determine the exact spatial position of the particles at the instance in time when the camera switches from one frame to the next. Moreover, the visible oscillations of the particles allow one to subdivide this time scale even further to obtain an idea about the spatial position of the particles after each acoustic cycle. Thus, for trajectories with good photographic resolution and sufficient oscillatory particle displacement (such as those in Fig. 4), one can construct a two-dimensional xy frame that identifies the instantaneous positions of both interacting particles in intervals of $T = 1/f$. In our example, the spatial resolution is given by $T = 1/900\text{ Hz}$, i.e., in time steps of 1.1 ms . Figure 4 combines the captured trajectories with a reconstruction frame indicating (with white dots) 53 instantaneous spatial positions for each of the two particles. From this, the vertical particle velocities can be extracted by determining first the distance between two neighboring white dots in the y direction and then relating it to the corresponding time period of 1.1 ms . As functions of time, Fig. 5(a) and (b) plot 52 instantaneous vertical particle velocities for the left and right particle, respectively. Digitalization of the image by the CCD camera restricts the maximum optical resolution to $3.7\text{ }\mu\text{m/pixel}$, leading to the observed “quantization” of the vertical particle velocities at two discrete values (23.31 and 26.64 mm/s , corresponding to 7 or 8 pixels/cycle). Nevertheless, it is clear from Figs. 5a and 5b that the vertical velocities stay nearly constant during the approach phase, an important result for the evaluation of other trajectories which are less favorably resolved than those shown in Fig. 4. Averaging over the 52 data points gives mean vertical particle velocities of 24.17 and 24.23 mm/s for the left and right particle, respectively, indicating an almost completely homogeneous motion of the two in the vertical direction. Moreover, these velocities compare well with the velocity determined from the trajectory length captured in the second video frame (enclosed by the two black cutlines in Fig. 4). Here the particles move over a distance of $784.4\text{ }\mu\text{m}$ (212 pixels) during $1/30\text{ s}$, resulting in a vertical particle velocity

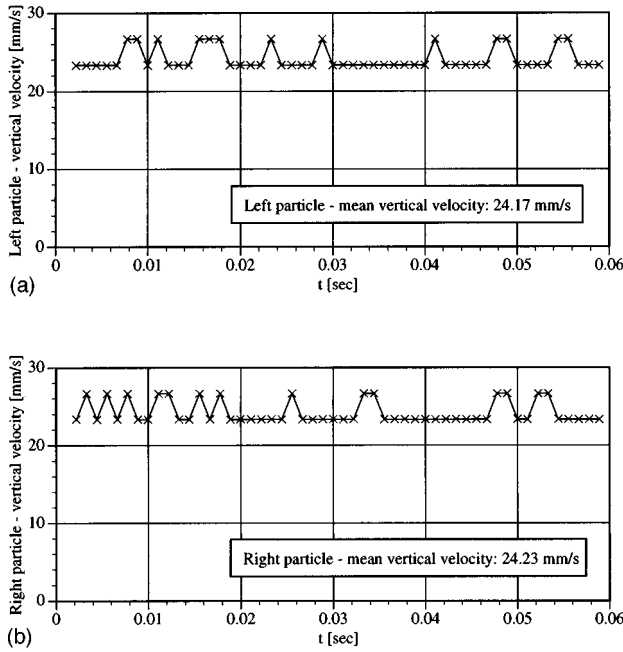


FIG. 5. Instantaneous vertical velocities of the left and right particle during the approach phase.

of 23.52 mm/s. This value stays within 3% of the velocities determined from single acoustic cycle evaluation. Thus, if poorer optical image quality prohibits the determination of the instantaneous velocities, the overall vertical particle motion can be approximated through evaluation of the trajectory length within a single video frame.

In Ref. 1 a qualitative comparison of the tuning fork agglomerations with existing agglomeration theories made clear that both mutual scattering interaction and mutual radiation pressure interaction⁹ predict, in contradiction to the observed attraction patterns, repulsion between particles aligned along the acoustic axis. However, a third theory, the acoustic wake effect model, implemented the identified feature of on-axis particle attraction. As a hypothesis, the observed tuning forks were then related to on-axis particle attraction due to acoustic wake interaction, a hydrodynamic effect caused by asymmetric flow fields which develop around the particles even at Reynolds numbers smaller than unity. This flow asymmetry leads to a region of lower pressure on the aft-side of a moving particle, causing nearby secondary particles to be dragged toward the primary one.¹² A theory to describe this effect based on Oseen flow fields¹⁴ was first proposed by Pshenai-Severin¹⁵ for equally sized particles, and then extended by Dianov *et al.*² to include interactions between differently sized particles.¹⁶ As a result, the theories determine analytical expressions for the relative time-averaged particle approach velocity \bar{u}_{12} due to the acoustic wake effect. It is important to point out that both models are limited to the on-axis case, i.e., to particle alignment that parallels the acoustic field vector. Note that, by contrast, the experimentally observed trajectories are in general imperfectly aligned with the acoustic axis.

We will now present an analysis which compares the observed tuning fork agglomerations with Dianov *et al.*'s model for the acoustic wake effect. To do so, we first recall

Dianov *et al.*'s analytical expression for the relative, time-averaged approach velocity between two particles aligned on the acoustic field vector. Based on Oseen flow field approximations they obtain

$$\bar{u}_{12} = \frac{3}{2} \frac{U_0}{\pi r_0} (a_1 l_1 + a_2 l_2), \quad (1)$$

where U_0 is the acoustic fluid velocity amplitude, r_0 the particle separation distance, and a_1 and a_2 are the radii of particle 1 and particle 2, respectively. The particle slip coefficients l_i in the Oseen regime are given for the i th particle by

$$l_i = \frac{n_i}{\sqrt{1 + 2h_i n_i^2 + h_i^2 n_i^4}} \quad (2)$$

with the particle slip coefficients n_i in the Stokes regime

$$n_i = \frac{\omega \tau_i}{\sqrt{1 + (\omega \tau_i)^2}} \quad (3)$$

and the constants

$$h_i = \frac{9}{2} \frac{U_0}{\pi \omega a_i} \frac{\rho_0}{\rho_p}. \quad (4)$$

In the above equations, τ_i is the particle relaxation time of the i th particle and ρ_0 is the density of the fluid. The particle slip coefficients l_i are nonlinearly dependent on frequency f , particle radius a , and the acoustic velocity amplitude U_0 . Equation (1) can be used to determine the time of convergence t from the initial particle separation distance $r_0(t_1)$ to the later one $r_0(t_2)$, i.e.,

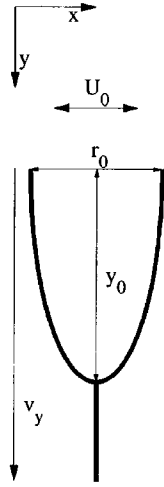
$$t = \frac{\pi}{3 U_0 (a_1 l_1 + a_2 l_2)} (r_0^2(t_1) - r_0^2(t_2)). \quad (5)$$

In our earlier paper¹ we integrated Eq. (1) numerically with a simple Euler's scheme (assuming a constant vertical particle velocity v_y) to determine the instantaneous spatial positions of an on-axis aligned, equally sized particle pair. This evaluation revealed tuning fork shaped particle attraction patterns which resembled those found in our experimental visualization experiments. In the present paper we employ the more accurate fourth-order Runge-Kutta method.¹⁸ Note that because of the smoothness of the trajectory function only insignificant differences were found between the results determined with the two different numerical techniques.

For comparison of the experimental results with the numerically determined tuning forks, a normalization is needed which allows for data interpretation independent of the vertical particle velocity v_y and the initial particle separation distance r_0 . To do so, we first note that the relative convergent velocity \bar{u}_{12} of Eq. (1) is inversely proportional to the particle separation distance r_0 . Therefore, we define the normalized separation distance $x_n = x/r_0$ so that all trajectories are initiated at a unit separation distance (see diagram in Table I).

Second, the length of the trajectory in the y direction is linearly dependent on the vertical particle velocity v_y . Therefore, division by v_y normalizes the experimental data

TABLE I. Evaluation of ten experimental tuning fork trajectories obtained at 800 Hz.

| Acoustic velocity amplitude U_0 [m/s] | Reynolds-number $Re = \frac{U_0 D}{\nu}$ | Vertical particle velocity v_y [mm/s] | Initial separation distance r_0 [μm] | Approach ratio $\frac{y_0}{r_0}$ | Normalized approach ratio y_{0n} | $v_t = 4.64$ mm/s $a = 4.05$ μm $r_{0n} = 1$ |
|--|---|--|--|-------------------------------------|---------------------------------------|---|
| 0.64 | 0.34 | 16.7 | 150 | 1.03 | 0.0077 |  |
| | | 15.5 | 194 | 1.44 | 0.0090 | |
| | | 13.1 | 237 | 1.83 | 0.0111 | |
| | | 17.7 | 150 | 1.73 | 0.0122 | |
| | | 15.0 | 140 | 1.68 | 0.0150 | |
| Mean value | | 15.6 | 174 | 1.54 | 0.0110 | |
| Dianov <i>et al.</i>'s theory | | | | | 0.0042 | |
| 0.81 | 0.43 | 21.8 | 206 | 1.12 | 0.0047 | |
| | | 15.8 | 193 | 1.16 | 0.0071 | |
| | | 14.2 | 160 | 1.01 | 0.0084 | |
| | | 18.0 | 150 | 1.71 | 0.0119 | |
| | | 10.9 | 206 | 1.44 | 0.0120 | |
| Mean value | | 16.1 | 183 | 1.29 | 0.0088 | |
| Dianov <i>et al.</i>'s theory | | | | | 0.003 | |

with respect to the vertical particle motion, and multiplication by a reference velocity v_t makes it dimensionless (v_t is chosen to be the theoretically determined settling velocity of a single particle, with $v_t = g \cdot \tau = 4.64$ mm/s for an $8.1\text{-}\mu\text{m}$ particle with $\rho_p = 2400$ kg/m³, $g = 9.81$ m/s², and τ is the relaxation time as defined above). Note that the experimentally observed vertical velocities v_y are determined by the gravitational bulk motion of a particle cloud and are therefore higher than v_t . This bulk motion originates from the noncontinuous particle feeding procedure employed; activated by mechanical vibrations, a large quantity of particles is suddenly released into the observation chamber, generating a particle cloud whose vertical velocity exceeds that of an individual particle (for details see Ref. 19). As pointed out in the discussion of Fig. 4, this bulk velocity remains constant for the duration of a single tuning fork approach. However, for different experiments the bulk velocities are found to vary in the range from 10.9 to 23.6 mm/s because of bulk density fluctuations. Therefore, it is necessary to normalize the y axis with regard to the bulk velocity. The suggested multiplication by v_t/v_y implies that all particles drop at a normalized gravitational velocity, i.e., the theoretical terminal settling velocity of an individual $8.1\text{-}\mu\text{m}$ particle.

Third, Eq. (5) shows that the total time t_0 needed for the two particles to collide is proportional to the square of the initial separation distance r_0 , i.e., $t_0 \sim r_0^2$ [collision condition with $r_0(t_1) = r_0$ and $r_0(t_2) = 0$]. Assuming a constant vertical particle velocity v_y , this proportionality is also valid with respect to the total length (or “collision distance”) of the tuning fork y_0 , namely $y_0 \sim r_0^2$ (because of $y_0 = v_y \cdot t_0$; see diagram in Table I). Thus, to make trajectories that originate from different initial separation distances comparable in terms of their y components, the model based on Eq. (5) suggests a division of the ordinate by r_0^2 . Finally, to bring the normalized y axis into a nondimensional form, a character-

istic length has to be multiplied, here chosen to be the particle radius $a = 4.05$ μm . As a result, the normalized and nondimensionalized vertical component of the trajectory takes the form $y_n = y \cdot (a/r_0^2 \cdot v_t/v_y)$. Knowing the initial separation distance and the vertical velocity of the particles, it is now possible to compare the data of different experiments and contrast them with numerical computations. This representation makes it possible to limit the evaluation of the particles convergence velocity to only one parameter, the acoustic velocity amplitude U_0 .

Because we obtained most of our experimental data with $8.1\text{-}\mu\text{m}$ microspheres at a frequency of 800 Hz, we will focus the subsequent analysis on the evaluation of these specific results. Table I summarizes some important data extracted from the analysis of ten different tuning fork trajectories. For each experiment, the table indicates in the first two columns the acoustic velocity amplitude U_0 (0.64 or 0.81 m/s) and the resulting Reynolds number $Re = U_0 D/\nu$ (0.34 or 0.43, respectively). Columns three and four give the vertical particle velocity v_y (ranging from 10.9 to 21.8 mm/s) and the initial particle separation distance r_0 (ranging from 140 to 237 μm). The particle approach ratio, which relates the total length of the tuning fork y_0 to the initial separation distance r_0 , is listed in column 5. Finally, the most meaningful data of the table appears in column 6 which presents, in ascending order, the normalized approach ratio $y_{0n}/r_{0n} = y_{0n}/1$. These values, which are graphically reproduced in Fig. 6, represent the normalized length/width ratios of the trajectories unbiased by the vertical particle velocity and the initial separation distance. Using this tool we will now compare ten experimental tuning forks and evaluate their dependence on the acoustic velocity excitation.

The mean values of y_{0n} at the two different acoustic velocities (Fig. 6) indicate slightly shorter trajectories for the higher acoustic excitation level (0.0110 at 0.64 m/s and

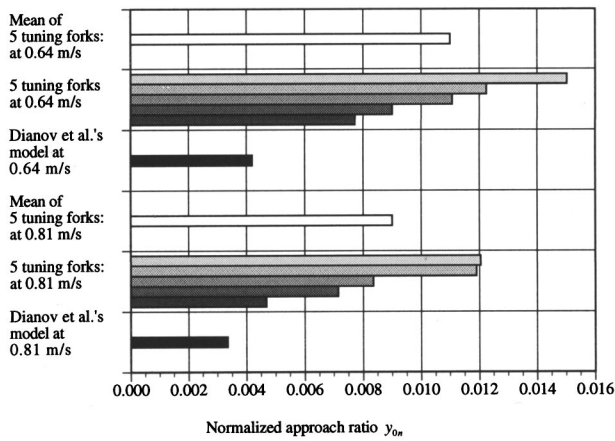


FIG. 6. Normalized approach ratios at two acoustic velocity amplitudes.

0.0088 at 0.81 m/s). This result remains consistent if we consider the intervals that the y_{0n} values span within each set of five experiments for identical acoustic excitation. At the smaller velocity 0.64 m/s we find y_{0n} to range between 0.0077 and 0.0150 (longer tuning forks), whereas at 0.81 m/s this range is 0.0047–0.0120 (shorter tuning forks). These results indicate that an increased acoustic excitation velocity leads to stronger particle interaction and higher convergent particle velocities. However, the large deviations of the y_{0n} values from the mean make it difficult to define a qualitative relationship between \bar{u}_{12} and U_0 . It is reasonable to assume that the smallest approach ratios found (i.e., 0.0077 and 0.0047) refer to geometrical particle configurations which more or less parallel the acoustic axis. Therefore, these values should be employed for comparison to the numerical simulation which is limited to the on-axis case. On the other hand, the more elongated tuning fork trajectories most likely originate from interacting particle pairs whose connecting axes form angles with respect to the acoustic vector. As detailed in Ref. 19, the visualization results give only limited information about this angle due to the loss of information by projecting a three-dimensional geometrical arrangement onto a 2D visualization image. A pair of particles may appear to be perfectly aligned in an image, but move at a certain angle with respect to the normal of the visualized image plane.

Based on the Runge–Kutta scheme, numerical tuning fork trajectories were generated for the two acoustic velocities employed in the experiments. After normalization, these trajectories exhibit approach ratios y_{0n} of 0.0042 and 0.0033 for the acoustic amplitudes 0.64 and 0.81 m/s, respectively (see Fig. 6). Hence, the shortest experimental trajectories, with $y_{0n}=0.0077$ at 0.64 m/s and $y_{0n}=0.0047$ at 0.81 m/s, exceed the numerical approach ratios by the factors 1.8 (0.64 m/s) and 1.4 (0.81 m/s). At first sight, this seems to indicate that Dianov *et al.*'s model overestimates the relative convergent velocity between the particles. However, as pointed out above, even for the shortest observed trajectories we cannot be certain that the particles are perfectly aligned with the acoustic field vector. For particle pairs interacting at an angle to the acoustic vector one expects, in accordance with our experimental results, less interaction and a reduced

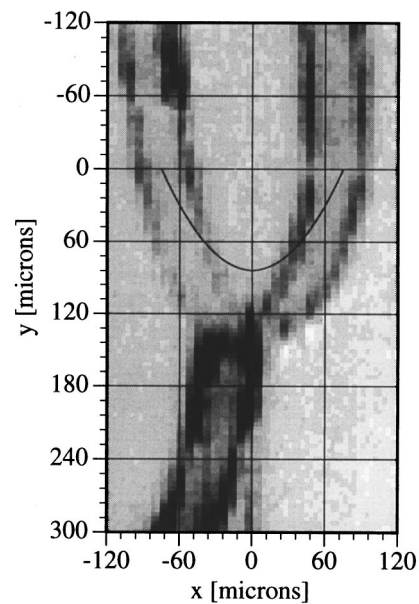


FIG. 7. Comparison of an experimental and a theoretical tuning fork trajectory of two 8.1- μm microspheres dropping vertically (particle density 2400 kg/m³—soda lime glass); the acoustic sound field vector is aligned horizontally; acoustic velocity amplitude 0.64 m/s; frequency 800 Hz; vertical particle velocity 16.7 mm/s; initial separation distance 150 μm .

convergent velocity.¹³ Moreover, Table I indicates that the Reynolds numbers of our experiments were relatively low (0.34 at 0.64 m/s and 0.43 at 0.81 m/s), whereas Dianov *et al.* base their model on Oseen flow approximations aimed at Reynolds numbers up to unity. This implies that the flow field asymmetries of the model as well as the resulting hydrodynamic interaction effects are more pronounced than those expected under the given experimental condition. Finally, the theory restricts the applicability of Eq. (1) to separation distances $r > \nu/U_0$. In our case this means that only trajectories with separation distances larger than 24 μm (at 0.64 m/s) and 19 μm (at 0.81 m/s) meet this criterion. For smaller separation distances down to the collision point the model serves only as a “good guess.” In the light of these limitations and uncertainties, it is almost surprising that the experimental and theoretical approach ratios differ only by a factor 1.4 at 0.81 m/s (and 1.8 at 0.64 m/s).

To finalize the analysis of the acoustic wake effect we will try to give the reader a visual impression of how the experimental tuning forks compare to Dianov *et al.*'s numerical trajectories. Figure 7 depicts a magnified image of the shortest observed tuning fork at 0.64 m/s ($y_{0n}=0.0077$) together with the corresponding numerical trajectory ($y_{0n}=0.0042$) starting at a separation distance of 150 μm and moving with a vertical velocity of 16.7 mm/s. The figure clearly illustrates that, even for this experimental tuning fork, the model overestimates the convergent velocities between the particles. As a result, the numerical calculations reveal a stronger curvature of the trajectories and show particles that collide within a shorter period of time.

We conclude that Dianov *et al.*'s acoustic wake effect model serves to roughly determine the convergent velocity between two particles with perfect on-axis alignment. For particles with angled spatial orientation, however, the experi-

mental approach rates are found to exceed the numerical predictions by a factor of 2 or 3. This makes clear that, while conceptually helpful, the model suffers from limitations due to its one-dimensional and approximate nature. In a similar manner, the experimental data are not quite sufficient to explicitly determine the actual spatial orientation of each observed tuning fork. Thus, at this stage of the investigation the experimental evaluation of the theoretical findings remains qualitative. Future work will concentrate on an extended 3-D acoustic wake effect model as well as on improved trajectory visualization techniques.

E. Captive force field hypothesis

The above analysis made clear that Dianov *et al.*'s acoustic wake effect model tends to overestimate the convergent velocity for on-axis aligned particles. This fact and the theoretical restrictions toward small inter-particle separation distances limit the applicability of the acoustic wake effect model to assess the captive force field hypothesis presented in our earlier paper.¹ There it was proposed that repulsive on-axis forces generated by the mutual radiation pressure interaction⁹ counterbalance the attractive motion of the acoustic wake effect, leading to stable geometrical arrangements of the interacting particles without being in physical contact with one another. These so-called "pseudoagglomerates" would then move jointly in the acoustic field maintained by the balance of forces at a determined inter-particle separation distance. Figure 13 of Ref. 1 showed that similar phenomena were indeed observed in our experiments at 600 Hz. Driven by the acoustic wake effect, the particles first merge into the typical tuning fork patterns. At a certain separation distance, however, the attractive motion comes to an end, leaving a narrowly spaced particle group which travels in parallel trajectories. Despite this experimental evidence, the quantitative evaluation of the phenomenon is hindered by the abovementioned restrictions of Dianov *et al.*'s model as well as by the fact that Song's mutual radiation pressure interaction model⁹ requires a very precise knowledge of the particle sizes. The manufacturer specifications of the employed glass microspheres were not sufficient to deliver the necessary accuracy. Thus, at this stage the captive force model can only be proposed but not be quantified with existing models.

II. CONCLUSIONS

The objective of the present paper was to evaluate existing acoustic agglomeration theories based on experimental findings presented in our earlier paper.¹ This comparison has led to a number of important conclusions:

(1) Measured acoustic particle entrainment rates were found to be generally lower than those predicted by the theory based on the BBO equation. This result is surprising because that a prior experimental study by Gucker and Doyle⁴ revealed entrainment rates that coincided with that theory. Note, however, that unpublished experimental results by González *et al.*⁷ confirm the lower entrainment rates found in the present investigation.

(2) The classical orthokinetic acoustic agglomeration hypothesis builds upon the relative entrainment motion of differently sized particles. In our experiments we were able to visualize typical orthokinetic entrainment situations in a polydisperse aerosol. However, we were unable to detect an incidence of particle collision or agglomeration originating from orthokinetic particle motion. Even though our experiments were limited to the low frequency range, these results give a first indication that other-than-orthokinetic effects may be determining the acoustic agglomeration process. Further experiments at higher frequencies are recommended.

(3) Song^{8,9} proposed that mutual scattering interactions cause attraction and collision between particles with off-axis orientation and repulsion in the case of on-axis alignment. These predictions were found to be in clear contradiction to the observed attraction patterns of particle pairs aligned on the acoustic field vector (tuning fork agglomerations).

(4) Tuning fork agglomerations were shown to display all the important features of acoustic wake particle interaction, a hydrodynamic effect based on asymmetric flow fields around the particles. To evaluate the observed tuning fork trajectories quantitatively we employed Dianov *et al.*'s² acoustic wake effect model based on Oseen flow fields. Because of its one-dimensional and approximate nature, however, the model was only applicable to particle pairs with perfect on-axis alignment. On the other hand, the actual alignment angle of a particle pair in the experiments was not precisely known. Despite these limitations of both the model and the experimental results, we were able to carry out a comparison by applying an appropriate normalization to the particle approach ratio. It was found that the model tended to overestimate the convergent particle velocity by a factor of at least 1.4. Thus, Dianov *et al.*'s model gave only a rough estimate of the maximum expected particle approach velocity. Future work will have to concentrate on a better 3-D acoustic wake effect model as well as on improved trajectory visualization techniques.

(5) Some of the observed tuning fork trajectories indicated that there may be a counterbalance of opposing forces in the close vicinity of the particles. This "captive force field," possibly arising from mutual radiation pressure repulsion and acoustic wake attraction, leads to groups of particles—called pseudoagglomerates—which travel jointly without being in actual physical contact with one another. At this stage, this hypothesis cannot be quantified due to inherent limitations of the existing models.

ACKNOWLEDGMENTS

This research was financed by Subcontract No. 10038-01 of the Manufacturing and Technology Conversion International, Inc. and by Contract No. DE-AC21-89MC26288 of the Department of Energy. Support from these institutions is gratefully acknowledged. In addition, we wish to express our appreciation to Dr. Iman, manager of the Mechanical Systems Dynamics Project at GE Corporate Research and Development Schenectady, New York, for providing us with experimental equipment. The experimental work would have been impossible without the help and equipment of Professor Santoro and his student Peter

Strakey, whom we sincerely thank for their assistance. In addition, we would like to thank Professor Temkin from Rutgers University, as well as Professor Gallego, Dr. Riera, and Itziar González for their assistance to this work at the Instituto de Acústica in Madrid, Spain. Last but not least, many thanks to David S. Bright (a subscriber to the NIH image newsgroup from the National Institute of Standards & Technology in Gaithersburg) for his help in processing some of this paper's digital images.

- ¹T. L. Hoffmann and G. H. Koopmann, "Visualization of acoustic particle interaction and agglomeration: Theory and experiments," *J. Acoust. Soc. Am.* **99**, 2130–2141 (1996).
- ²D. V. Dianov, A. A. Podol'skii, and V. I. Turubarov, "Calculation of the hydrodynamic interaction of aerosol particles in a sound field under Oseen flow conditions," *Sov. Phys. Acoust.* **13**, 314–319 (1968).
- ³W. König, "Hydrodynamisch-akustische Untersuchungen: I. Über das Mitschwingen einer Kugel in einer schwingenden Flüssigkeit," *Ann. Phys. Chem.* **42**, 352–370 (1891).
- ⁴F. T. Gucker and G. J. Doyle, "The amplitude of vibration of aerosol droplets in a sonic field," *J. Phys. Chem.* **60**, 989–996 (1956).
- ⁵S. Temkin, *Elements of Acoustics* (Wiley, New York, 1981).
- ⁶S. Temkin and C.-M. Leung, "On the velocity of a rigid sphere in a sound wave," *J. Sound Vib.* **49**, 75–92 (1976).
- ⁷I. González, T. L. Hoffmann, and J. A. Gallego-Juárez, "Measurement of acoustic particle entrainment factors," *J. Aerosol Sci.* (in preparation).
- ⁸L. Song, "Modeling of acoustic agglomeration of aerosol particles," Ph.D. dissertation, Pennsylvania State University (1990).
- ⁹L. Song, G. H. Koopmann, and T. L. Hoffmann, "An improved theoretical model of acoustic agglomeration," *Trans. ASME* **116**, 208–214 (April 1994).
- ¹⁰In Section 4.10 (pp. 240–244) on "Oseen's improvement of the equation of flow due to moving bodies at small Reynolds number," Batchelor's (Ref. 11) standard treatise on fluid dynamics refers to the flow region directly behind the sphere as a 'wake' (pp. 242–243 and Figure 4.10.1). In this paper we thus adopt Batchelor's use of the expression 'wake' for the relatively small Reynolds numbers that correspond to the Oseen flow regime.
- ¹¹G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge U.P., Cambridge, 1987).
- ¹²Temkin and Ecker (Ref. 13) (p. 468) identify the reason for an observed attraction between two droplets as "... the transient wake that is produced in the lee side of one droplet, due to the passage of the wave. If a second droplet is in the vicinity of this wake, it will experience smaller fluid forces than the first. The first droplet may therefore move much more rapidly than the second, possibly producing collision."
- ¹³S. Temkin and G. Z. Ecker, "Droplet pair interactions in a shock-wave flow field," *J. Fluid Mech.* **202**, 467–497 (1989).
- ¹⁴Batchelor (Ref. 11) (p. 256) discusses the physics of low Reynolds number flow fields around particles: "At Reynolds numbers small compared with unity, the dominant process in the flow is the diffusion of vorticity away from the body.... The Oseen improved approximation of §4.10 takes account of inertia forces partially, and the vorticity here diffuses from a source which is moving steadily.... Thus, for $R \ll 1$ [R = Reynolds number], when Stokes and Oseen approximations are applicable, the flow has fore-and-aft symmetry near the body with that same symmetry, but distinct asymmetry further out."
- ¹⁵S. V. Pshenai-Severin, "On the convergence of aerosol particles in a sound field under the action of the Oseen hydrodynamic forces," *Dokl. Akad. Nauk SSSR* **125**, 775–778 (1959).
- ¹⁶Even though it is not the intention of this paper to prove the different theories for correctness, it should be noted here that both Pshenai-Severin's and Dianov *et al.*'s theory are based on the superposition of the two particles' individual flow fields. Because of this approximation, neither of the theories strictly satisfies the boundary conditions on the particles' surfaces [see discussion in Fuchs (Ref. 17) pp. 46–47, pp. 101–102, and p. 328].
- ¹⁷N. A. Fuchs, *The Mechanics of Aerosols* (Pergamon, Oxford, 1964).
- ¹⁸W. H. Press, B. P. Flannery, S. A. Teukolsky, and V. T. William, *Numerical Recipes: The Art of Scientific Computing* (Cambridge U.P., New York, 1986).
- ¹⁹T. L. Hoffmann and G. H. Koopmann, "A new technique for visualization of acoustic particle agglomeration," *Rev. Sci. Instrum.* **65**, 1527–1536 (1994).

Vibration and wave localization in a nearly periodic beaded string

Gisli Óttarsson

Mechanical Dynamics, Inc., 2301 Commonwealth Boulevard, Ann Arbor, Michigan 48105

Christophe Pierre

Department of Mechanical Engineering and Applied Mechanics, The University of Michigan, Ann Arbor, Michigan 48109-2125

(Received 6 April 1995; accepted for publication 22 February 1997)

The dynamics of an exact, linear model of a string with attached beads are examined. A study of the perfectly periodic system is performed using a transfer matrix formulation. A closed-form solution for the natural frequencies of the finite system is obtained. The relationship between natural frequencies and the passbands and stop bands of propagating waves is studied. The effect of random disorder of (a) bead spacing and (b) bead mass is examined and interesting, fundamental differences are observed. In the case of bead-spacing disorder both weak and strong localization occur, whereas bead-mass disorder only causes weak localization. *Localization*, the spatial confinement of vibration energy due to periodicity breaking disorder, is quantified by the *localization factor*. Analytical approximations of the localization factor are derived, in the limits of large and small coupling to disorder ratios. The results are verified through Monte Carlo simulations. The natural frequencies and modes of the disordered system are examined. A dispersion of the natural frequencies is evidenced, presented as a curve veering phenomenon. The existence of localized mode shapes is demonstrated. © 1997 Acoustical Society of America. [S0001-4966(97)04406-8]

PACS numbers: 43.40.Cw [CBB]

INTRODUCTION

The loaded string, a taut string with equally spaced attached beads, has long been a topic of interest in the field of vibrations. Among those who have considered this problem are Lagrange, Rayleigh, and Routh. For a brief overview of the early history of this problem, the readers are referred to Ref. 1.

The periodic beaded string is a member of a class of systems referred to as periodic, monocoupled systems, consisting of a number of identical subelements connected in an identical manner through one coupling coordinate. Much work has been performed in the area of periodic systems, leading to an understanding of how wave carrying motion can only occur in distinct frequency bands known as *propagation zones* or *passbands*. For a full discussion of wave propagation in periodic systems see Ref. 2.

More recently research has focused on the effect of disorder in otherwise periodic systems. In the loaded string, disorder may occur when either the spacing of the attached beads or the bead mass is slightly disordered. Disorder has been shown to cause “localization,” a phenomenon which may be expected to occur in any nearly periodic structure where perfect periodicity is prevented by manufacturing tolerances and other defects. Localization, like damping, causes spatial decay of the vibration amplitude along the structure, but for vastly different reasons. In the case of damping, energy is dissipated from the system, whereas in the case of localization the energy is merely confined to the neighborhood of the source of excitation. Localization occurs because waves propagating away from the energy source are reflected at the interfaces of the subtly different subsystems making up

the nearly periodic structure. This energy localization leads to much higher amplitudes than would be predicted if perfect periodicity were assumed, with possibly disastrous effects, e.g., in bladed disk assemblies.

Examples of structures that have been analyzed for disorder effects include blade assemblies,^{3,4} multispan structures,^{5–7} and some large space structures.⁸ Kissel⁹ studied localization in several infinite one-dimensional systems with a wave-propagation approach, applying Furstenberg’s theorem on the limiting products of random transfer matrices. Limitations of the systems he examined restricted his findings mostly to weak localization, less important than strong localization in typical engineering applications.

The taut string with attached beads is one (nearly) periodic structure that has yet to receive the attention it deserves, in spite of having some particular features that give new insight into the localization phenomenon. The loaded-string system has characteristics that distinguish it substantially from the other systems discussed in Refs. 3–9. First of all, the lack of immovable points in the sub-elements allows global motion of the assembly. The coupling between bays is also different, as it is inertial in nature rather than due to the stiffness. This is because the string tension, contrary to intuition, cancels out of the coupling parameter used for the system. The authors are only aware of a few papers on the topic. Hodges and Woodhouse¹⁰ used a pendulum-chain analogy to perform a statistical perturbation analysis of localization in a loaded string. Evidence of strong and weak localization was found, in the cases of weak and strong bay-to-bay coupling, respectively, but a systematic parameter study was not performed, perhaps due to limitations of the pendulum analogy. He and Maynard¹¹ performed an experimental investigation

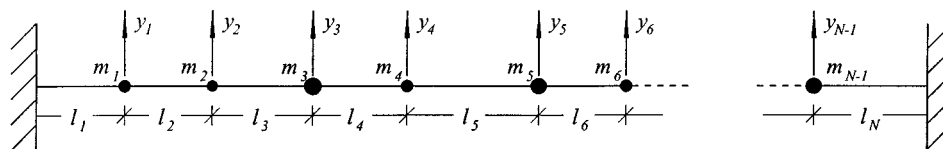


FIG. 1. A nearly periodic string with attached beads.

of localization using a nearly periodic loaded string. Their findings, which include well-defined cases of strong and weak localization, provide experimental support for many of the conclusions of the current theoretical investigation. More recently Rajagopal¹² calculated localized modes of vibration in a seven-bead nearly periodic string and verified their occurrence by an experiment. A disorder-induced widening of the natural frequency clusters was evidenced. Finally, note that in an 1981 paper Sergev and Iwan¹³ studied the mode shapes of a string with beads attached at different intervals. One of the figures in their paper illustrates a mode shape which in the words of the authors "... is complex and much beyond the range of intuition." The mode in question is in fact highly localized. The authors appear to have been unfamiliar with the localization effect not widely recognized at the time in the field of structural dynamics.

In the first part of the paper the equations of motion of the beaded string are derived, from wave and vibration perspectives. In the second part the ordered system dynamics are examined extensively. The nature of the passband-stopband structure is explained and the free vibration eigenvalue problem of the ordered system is solved analytically using a wave propagation approach. The results are compared with Mead's survey of moncoupled periodic systems.² The third part deals with the disordered system. Techniques used to calculate the localization factor are reviewed and utilized to derive analytical expressions of it in the two limiting cases of weak and strong coupling. The analytical expressions are verified against the results of Monte Carlo simulations. Finally, the results of a numerical analysis of the natural frequencies and modes of a disordered system are presented.

I. FORMULATION

The system under investigation is the undamped, taut string with attached beads shown in Fig. 1. The system is considered almost periodic with two possible types of disorder, namely in the *mass* as well as the *spacing* of the attached beads. Either disorder is assumed to be an independent, identically (and uniformly) distributed random variable of mean and standard deviation $(m, m\sigma_e)$ and $(l, l\sigma_\delta)$, respectively. The mass per unit length of the string ρ and the tension T are both considered constant and the bending stiffness of the string is ignored. In order to analyze the motion of the string two methods of formulation are possible, as outlined in the following sections. The former uses transfer matrices to study the propagation of incident waves in an infinite beaded string, whereas the latter considers the steady-state response of a finite system harmonically driven at one end.

A. Transfer matrix formulation

Any (nearly) periodic structure may be seen as an assembly of *bays*. A bay is defined as exactly one spatial period in the periodic structure, such that only bays that are nearest neighbors are coupled. A transfer matrix modeling of such a structure requires the definition of a *state vector*, which describes either the state of a bay or the state at the interface between bays. The dimension of the state vector may be shown to be twice the number of coupling coordinates between the bays.² To a state vector corresponds a transfer matrix that relates the states of adjacent bays or the states at adjacent interfaces.

Figure 2 depicts one possible definition of a bay for the loaded string. The bay contains one whole bead and one whole string segment. Other bay definitions are possible, as shown below.

As the deflection and slope of a string segment are dependent variables, adjacent bays in the string have only one coordinate in common. Hence the system is moncoupled and a state vector for the system has dimension two. One possible state vector describes the displacement and slope at the right bay interface with a corresponding transfer matrix equation, relating the states at the *interfaces* on each side of bay *i*:

$$\begin{bmatrix} y \\ \phi^R \end{bmatrix}_i = \mathbf{T}_i \begin{bmatrix} y \\ \phi^R \end{bmatrix}_{i-1}, \quad i = 1, \dots, N. \quad (1)$$

Another possible choice of a state vectors is to describe the displacement at the two ends of the bay, with a corresponding transfer matrix relating the state of bays *i* and *i-1*:

$$\begin{bmatrix} y_i \\ y_{i-1} \end{bmatrix} = \mathbf{Z}_{i,i-1} \begin{bmatrix} y_{i-1} \\ y_{i-2} \end{bmatrix}, \quad i = 1, \dots, N. \quad (2)$$

Both the above state vector definitions will be useful in this study. From the wave equation

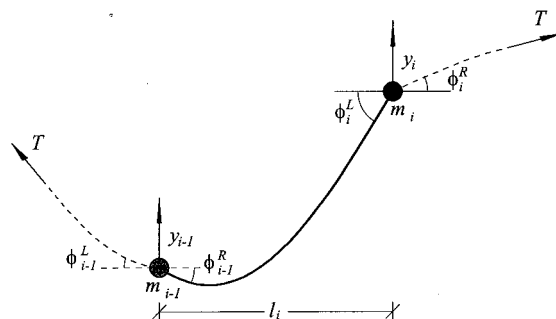


FIG. 2. A typical bay element.

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}, \quad \text{where} \quad c^2 = \frac{T}{\rho}, \quad (3)$$

one obtains a field transfer matrix connecting the states at the ends of a string segment of length l_i ,

$$\mathbf{T}_{\text{string}}(l_i) = \begin{bmatrix} \cos \frac{\omega l_i}{c} & \frac{c}{\omega} \sin \frac{\omega l_i}{c} \\ -\frac{\omega}{c} \sin \frac{\omega l_i}{c} & \cos \frac{\omega l_i}{c} \end{bmatrix}, \quad (4)$$

where harmonic motion has been assumed. The equation of motion of an attached bead of mass m_i , yields a point transfer matrix relating the slope of the string on the two sides of the bead, ϕ_i^R and ϕ_i^L ,

$$\mathbf{T}_{\text{bead}}(m_i) = \begin{bmatrix} 1 & 0 \\ -\frac{m_i \omega^2}{T} & 1 \end{bmatrix}. \quad (5)$$

The transfer matrix of the entire bay is then the product of the two,

$$\begin{bmatrix} y \\ \phi^R \end{bmatrix}_i = \mathbf{T}_{\text{bead}}(m_i) \mathbf{T}_{\text{string}}(l_i) \begin{bmatrix} y \\ \phi^R \end{bmatrix}_{i-1} = \mathbf{T}_i \begin{bmatrix} y \\ \phi^R \end{bmatrix}_{i-1}. \quad (6)$$

The bead masses m_i and bay lengths l_i are treated as random variables

$$m_i = m(1 + \epsilon_i) \quad \text{and} \quad l_i = l(1 + \delta_i), \quad (7)$$

where l and m are average values of l_i and m_i , and δ_i and ϵ_i are disorder random variables with mean zero and variance σ_δ^2 and σ_ϵ^2 , respectively. Introducing the dimensionless quantities $\bar{\omega}_i$, \bar{R}_i , and \bar{y}_i such that

$$\bar{\omega}_i = \frac{\omega l_i}{c} = \bar{\omega}(1 + \delta_i), \quad \bar{R}_i = \frac{T l_i}{m_i c^2} = \frac{\rho l_i}{m_i} = \bar{R} \frac{(1 + \delta_i)}{(1 + \epsilon_i)}, \quad (8)$$

and

$$\bar{y}_i = \frac{y_i}{l},$$

where $\bar{\omega}$ and \bar{R} are average values of $\bar{\omega}_i$ and \bar{R}_i , respectively, leads to the transfer matrix \mathbf{T}_i in Eq. (1), as

$$\mathbf{T}_i = \begin{bmatrix} \cos \bar{\omega}_i & \frac{\sin \bar{\omega}_i}{\bar{\omega}} \\ -\bar{\omega} \left(\sin \bar{\omega}_i + \frac{\bar{\omega}}{\bar{R}_i} \cos \bar{\omega}_i \right) & \cos \bar{\omega}_i - \frac{\bar{\omega}_i}{\bar{R}_i} \sin \bar{\omega}_i \end{bmatrix}. \quad (9)$$

Here \mathbf{T}_i is a transfer matrix relating the states at interfaces bordering the i th bay. The dimensionless numbers $\bar{\omega}$ and \bar{R} will be referred to as *dimensionless frequency* and *dimensionless coupling*, respectively. In the case of large values of \bar{R} the bead mass is small compared with that of a single string segment, leading to strong coupling of the string, segments, and vice versa. It is noteworthy that the coupling of adjacent string segments only depends on inertia properties, not on the string tension. The dimensionless coupling will

prove to be an important parameter of the system. The interface state vector $[\bar{y}, \phi^R]_i^T$ and the bay state vector $[\bar{y}_i, \bar{y}_{i-1}]^T$ are related by the transformation

$$\begin{bmatrix} \bar{y}_i \\ \bar{y}_{i-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{\bar{\omega}_i \sin \bar{\omega}_i}{\bar{R}_i} - \cos \bar{\omega}_i & \frac{\sin \bar{\omega}_i}{\bar{\omega}} \end{bmatrix} \begin{bmatrix} \bar{y} \\ \phi^R \end{bmatrix}_i \\ = \mathbf{S}_i \begin{bmatrix} \bar{y} \\ \phi^R \end{bmatrix}_i. \quad (10)$$

The singular value $\bar{\omega}_i = n\pi$ is a natural frequency of the i th string segment, with fixed ends, and leads to $[\bar{y}_i, \bar{y}_{i-1}]^T = \mathbf{0}$. Hence the transfer matrix relating the states of the two adjacent bays i and $i-1$ is

$$\begin{bmatrix} \bar{y}_i \\ \bar{y}_{i-1} \end{bmatrix} = \mathbf{S}_i \mathbf{T}_i \mathbf{S}_{i-1}^{-1} \begin{bmatrix} \bar{y}_{i-1} \\ \bar{y}_{i-2} \end{bmatrix} = \begin{bmatrix} \beta_{i,i-1} & -\alpha_{i,i-1} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{y}_{i-1} \\ \bar{y}_{i-2} \end{bmatrix} \\ = \mathbf{Z}_{i,i-1} \begin{bmatrix} \bar{y}_{i-1} \\ \bar{y}_{i-2} \end{bmatrix}, \quad (11)$$

where

$$\beta_{i,i-1} = \frac{\sin(\bar{\omega}_i + \bar{\omega}_{i-1})}{\sin \bar{\omega}_{i-1}} - \frac{\bar{\omega}_{i-1}}{\bar{R}_{i-1}} \sin \bar{\omega}_i \quad (12)$$

and

$$\alpha_{i,i-1} = \frac{\sin \bar{\omega}_i}{\sin \bar{\omega}_{i-1}}. \quad (13)$$

It should be noted that the disadvantage of the displacement-only state variable representation is that the transfer matrix depends upon the characteristics of two adjacent bays, and hence is a function of both disorder random variables, δ_{i-1} and δ_i . The two random transfer matrices relating the states of three consecutive bays are no longer independent and averaging becomes difficult. For this reason, the displacement-only representation is less useful when analyzing disordered systems than the displacement/slope representation whose transfer matrix depends only on the parameters of a single bay.

It is noteworthy that the dynamics of the infinite-degree-of-freedom string has been described in an exact way using only the deflections of the coupling coordinates without requiring any kind of approximate discretization. This formulation is exact and the deflections of all points of bay i , $Y_i(x)$, may be recovered using the deflection of two adjacent points as

$$Y_i(x) = y_i \frac{\sin[\bar{\omega}_i x]}{\sin \bar{\omega}_i} + y_{i-1} \frac{\sin[\bar{\omega}_i(1-x)]}{\sin \bar{\omega}_i}, \quad (14)$$

$$0 \leq x \leq 1, \quad \bar{\omega}_i \neq n\pi.$$

B. Finite string dynamics

From Eq. (11) the deflections at three adjacent bays are related by:

$$y_i - \beta_{i,i-1} y_{i-1} + \alpha_{i,i-1} y_{i-2} = 0. \quad (15)$$

A finite $N+2$ bead system yields a system of N equations with $N+2$ unknowns. We apply boundary conditions corresponding to the first bead ($i=0$) being fixed and the last bead ($i=N+1$) being harmonically excited with amplitude Y and find a $N \times N$ tridiagonal system of equations,

$$\begin{bmatrix} \beta_{N+1,N} & -\alpha_{N+1,N} & 0 & \dots & 0 & 0 \\ -1 & \beta_{N,N-1} & -\alpha_{N,N-1} & \dots & 0 & 0 \\ 0 & -1 & \beta_{N-1,N-2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \beta_{3,2} & -\alpha_{3,2} \\ 0 & 0 & 0 & \dots & -1 & \beta_{2,1} \end{bmatrix} \times \begin{bmatrix} y_N \\ y_{N-1} \\ y_{N-2} \\ \vdots \\ y_2 \\ y_1 \end{bmatrix} = \begin{bmatrix} Y \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad (16)$$

which will be referred to as

$$\text{tridiag}[-1; \beta_{i,i-1}; -\alpha_{i,i-1}] \mathbf{y} = \mathbf{f}. \quad (17)$$

II. ORDERED SYSTEM DYNAMICS

In the ordered string all the string segments have equal length and all the beads have equal mass. We denote this with a subscript 0. It is advantageous to use the displacement representation of the state vector and Eq. (11) simplifies to

$$\begin{aligned} \begin{bmatrix} \bar{y}_i \\ y_{i-1} \end{bmatrix} &= \mathbf{S}_0 \mathbf{T}_0 \mathbf{S}_0^{-1} \begin{bmatrix} \bar{y}_{i-1} \\ \bar{y}_{i-2} \end{bmatrix} = \begin{bmatrix} \beta_0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{y}_{i-1} \\ \bar{y}_{i-2} \end{bmatrix} \\ &= \mathbf{Z}_0 \begin{bmatrix} \bar{y}_{i-1} \\ \bar{y}_{i-2} \end{bmatrix}, \end{aligned} \quad (18)$$

where

$$\beta_0 = 2 \cos \bar{\omega} - \frac{\bar{\omega}}{R} \sin \bar{\omega}. \quad (19)$$

A. Wave propagation in the ordered string

It has long been known that energy-carrying motions in periodic structures only occur in isolated frequency ranges known as *passbands*. Outside the passbands only attenuated standing waves (or, more rarely, complex waves) can take place.² A physical understanding of these wave-propagation characteristics can be gained through the diagonalization of Eq. (2). This requires the solution of the eigenvalue problem:

$$\begin{bmatrix} y_i \\ y_{i-1} \end{bmatrix} = \mathbf{Z}_0 \begin{bmatrix} y_{i-1} \\ y_{i-2} \end{bmatrix} = \lambda \begin{bmatrix} y_{i-1} \\ y_{i-2} \end{bmatrix}, \quad i = 1, \dots, N. \quad (20)$$

Equation (20) yields the eigenvalues and eigenvectors of \mathbf{Z}_0 . Since \mathbf{Z}_0 is real with $\det \mathbf{Z}_0 = 1$, the eigenvalues are reciprocal and are either real or complex conjugates. The eigenvectors of \mathbf{Z}_0 define wave modes, or characteristic waves, which propagate along the structure in such a way that the state vector is multiplied by a complex scalar λ as

the wave passes through each bay. The eigenvalues appear as a reciprocal pair,

$$\lambda, \lambda^{-1} = \frac{\beta_0}{2} \pm \sqrt{\left(\frac{\beta_0}{2}\right)^2 - 1}, \quad \lambda \in \mathbb{C}, \quad \beta_0 \in \mathbb{R}, \quad (21)$$

and define the frequency-dependent propagation properties of the corresponding wave modes. We choose the convention that λ has modulus greater than or equal to one. This associates λ with the wave mode traveling or attenuating in the direction of decreasing bay number. We shall call this the left direction. The wave modes appear as the eigenvectors, $[1, \lambda^{-1}]^T$ and $[1, \lambda]^T$, corresponding to the eigenvalues λ and λ^{-1} , respectively. The latter eigenvector could also be written as $[\lambda, 1]^T$, which shows that the eigenvectors are equivalent except for their direction of travel. This is reasonable since bays may be numbered in either the left or right directions with identical results. For a finite structure, the normal modes form a basis for all vibration shapes in a structure. Similarly, all possible waveforms in a structure may be written as a linear combination of the pair of wave modes.

The eigenvectors of \mathbf{Z}_0 are arranged as the columns of the matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ \lambda^{-1} & \lambda \end{bmatrix}, \quad (22)$$

which defines the transformation

$$\begin{bmatrix} y_i \\ y_{i-1} \end{bmatrix} = \mathbf{X} \begin{bmatrix} L \\ R \end{bmatrix}_i \quad (23)$$

from physical coordinates to left- and right-traveling wave coordinates at bay i , corresponding to the wave-mode basis. The displacement transfer matrix \mathbf{Z}_0 thus is transformed into a diagonal wave transfer matrix as

$$\begin{bmatrix} L \\ R \end{bmatrix}_i = \mathbf{X}^{-1} \mathbf{Z}_0 \mathbf{X} \begin{bmatrix} L \\ R \end{bmatrix}_{i-1} = \mathbf{W}_0 \begin{bmatrix} L \\ R \end{bmatrix}_{i-1} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix}_{i-1}. \quad (24)$$

Let $\lambda = e^{\mu}$ define the complex *propagation constant* μ with $\mu = \gamma + jk$. Here γ , the real part of the propagation constant, is the rate of exponential attenuation of the wave amplitude from one bay to the next. The imaginary part k is the *wave number*, the difference in phase between the motion of adjacent bays. The propagation constant contains all the information about the frequency-dependent propagation of waves through the assembly. Since β_0 in Eq. (19) is a real valued function of dimensionless frequency $\bar{\omega}$, we distinguish between the following cases:

$|\beta_0(\bar{\omega})| < 2$. In this case λ and λ^{-1} are complex conjugates of magnitude 1, yielding $\gamma = 0$. These frequencies define passbands in which waves travel without attenuation. From the real part of Eq. (21), the wave number is related to frequency by the dispersion relation:

$$2 \cos k = \beta_0(\bar{\omega}), \quad 0 < k < \pi. \quad (25)$$

For a given value of k , Eq. (25) has as many frequency solutions as there are degrees of freedom in each bay.¹⁴ Hence, the number of passbands equals the number of degrees of freedom in each bay and the beaded string has an

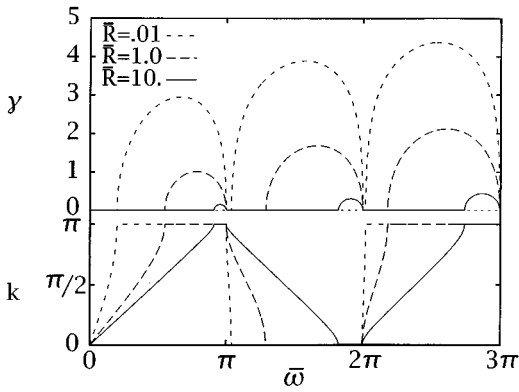


FIG. 3. Exponential decay γ and wave number k plotted as a function of dimensionless frequency $\bar{\omega}$ for various values of dimensionless coupling \bar{R} .

infinite number of passbands. One can reason that the natural frequencies of the system will lie in the passband because a perfectly periodic system must have periodic modes. The frequencies at which $k = \pi/2$ will be referred to as *midband frequencies*. Note that the midband frequency is not necessarily located close to the mean frequency in the passband.

$|\beta_0(\bar{\omega})| > 2$. The eigenvalues are real, hence γ is non-zero, leading to attenuation. Also, adjacent bays are vibrating either in phase or out of phase, $k=0$ or $k=\pi$, respectively, which implies that these are standing waves. These frequency ranges define *stop bands*.

$|\beta_0(\bar{\omega})| = 2$. This gives $\gamma=0$ and $k=0$ or $k=\pi$, and defines the bounding frequencies or the passband/stopband edges.

Figure 3 shows the exponential decay and wave number in the first three stopband/passband pairs for three systems with a range of coupling values. Apparent are the alternating stopbands ($\gamma \neq 0$ and $k=0$ or $k=\pi$) and passbands ($\gamma=0$ and $0 < k < \pi$). It should be noted that the first band is a passband, i.e., low-frequency waves travel without attenuation.

B. Properties of the bounding frequencies

It was shown above that $|\beta_0(\bar{\omega}, \bar{R})| = 2$ yields the frequencies separating stopbands and passbands. The equation

$$\beta_0 = 2 \cos \bar{\omega} - \frac{\bar{\omega}}{\bar{R}} \sin \bar{\omega} = \pm 2, \quad (26)$$

has two families of solutions corresponding to the upper and the lower passband edge, respectively.

The lower passband edge corresponds to the $\bar{\omega} = n\pi$ family of solutions shown to make the S_0 matrix singular. These frequencies are natural frequencies of the individual string segments with the attached beads coming to rest. This family of solution is obviously independent of the bead mass and hence independent of the coupling parameter \bar{R} .

The upper passband edge corresponds to the other less trivial family of solutions. The definition of a different bay type aids in the understanding of the behavior at these frequencies. The alternative bay element, illustrated in Fig. 4,

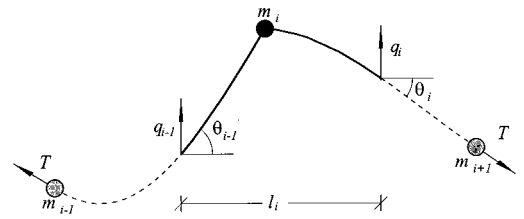


FIG. 4. An alternative bay with the bead in the middle.

has the bead at the center so that the state variables become $[q_i, \theta_i]^T$ by midstring. Using, the previously defined transfer matrices for a string segment and a point mass [Eqs. (4) and (5)], we obtain:

$$\begin{bmatrix} q \\ \theta \end{bmatrix}_i = \mathbf{T}_{\text{string}}\left(\frac{l}{2}\right) \mathbf{T}_{\text{bead}}(m) \mathbf{T}_{\text{string}}\left(\frac{l}{2}\right) \begin{bmatrix} q \\ \theta \end{bmatrix}_{i-1}, \quad (27)$$

$$\begin{bmatrix} \bar{q} \\ \theta \end{bmatrix}_i = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix} \begin{bmatrix} \bar{q} \\ \theta \end{bmatrix}_{i-1}, \quad (28)$$

where $\bar{q} = q/l$. This relationship between the midstring deflections and slopes may be rearranged to a form

$$\begin{bmatrix} \bar{q}_i \\ \bar{q}_{i-1} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \theta_i \\ \theta_{i-1} \end{bmatrix}. \quad (29)$$

The determinant of the matrix \mathbf{A} may be shown to be

$$\det[\mathbf{A}] = \frac{1}{\bar{\omega}^2} \frac{(1 - \cos \bar{\omega})}{(1 + \cos \bar{\omega})} \frac{(2 + \beta_0)}{(2 - \beta_0)}. \quad (30)$$

It follows that $\beta_0 = 2$ implies $\theta_i = 0$ and $\beta_0 = -2$ implies $q_i = 0$. In other words, an infinite, ordered string with attached beads vibrating at the frequency of the upper bound of a passband will either have zero displacement or zero slope at midbay. Furthermore, Eq. (30) shows that at the lower passband edge where the beads are at rest, the string center either has zero slope or zero displacement as expected for the sinusoidal shape.

Since the lower edge of the $(n+1)$ th passband ($n=0, \dots$) occurs at $\bar{\omega} = n\pi$, the passband width depends solely on the upper passband edge frequency, which is a function of both the interbay coupling \bar{R} and the passband number. Solving Eq. (26) in the limiting case when the ratio $\bar{\omega}/\bar{R}$ tends to infinity yields the two solutions $\bar{\omega}_{\text{left}} = n\pi$ and $\bar{\omega}_{\text{right}} \rightarrow n\pi$, so that for weak coupling and/or high passband number the passbands vanish. Recall that $\bar{\omega} = n\pi$ is the natural frequency of a string between fixed supports, i.e., for stationary beads. This is intuitively correct since the inertia of the beads will prevent them from participating in the motion of the string if their mass is high or if the motion is rapid. This was observed by Hodges and Woodhouse.¹⁰

Conversely, as $\bar{R} \rightarrow \infty$, that is the case of vanishing bead mass, the bays are perfectly coupled. In this case the stopbands vanish, and waves of all frequencies travel without attenuation.

Finally it may be shown that the maximum attenuation in the stopbands increases logarithmically with \bar{R} as well as the stop-band number, i.e., high-frequency waves are more strongly attenuated than low-frequency waves (see Fig. 3).

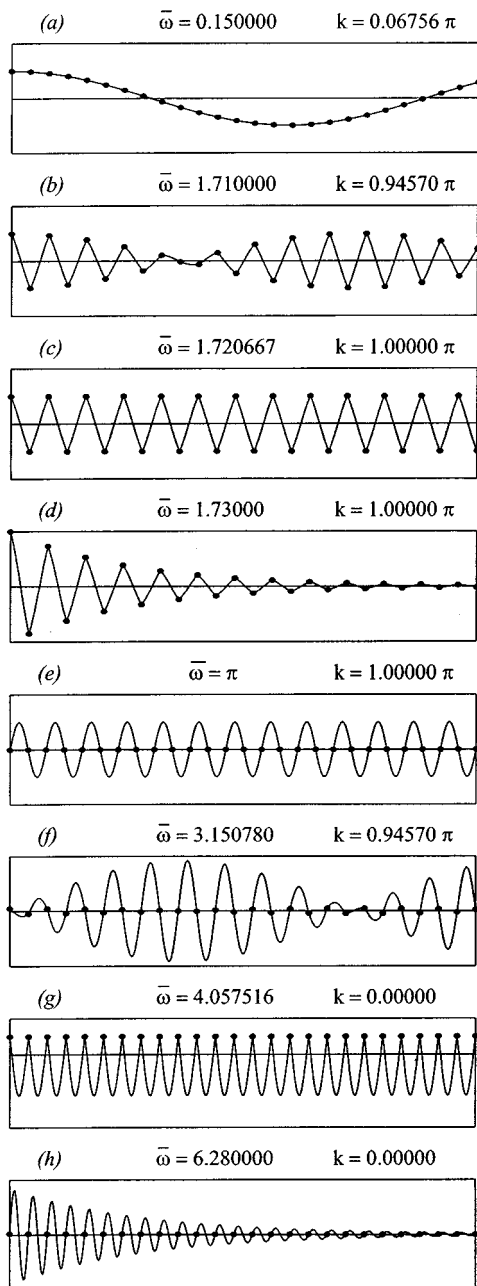


FIG. 5. Right traveling waves of amplitude 1.0 of various frequencies in an infinite string with $\bar{R}=1.0$. (a) An arbitrary traveling wave in the first passband. The frequency and wave number are very low and the mode is therefore highly global. It is of interest to note that in the first passband the frequency is lower than the first natural frequency of a fixed string of length l , which is $\bar{\omega}=\pi$. Therefore the string is participating very little in the vibration but is instead almost straight between the beads. (b) Another waveform in the first passband but with higher frequency and wave number. (c) A beaded string vibrating at the frequency of the upper edge of the first passband. As predicted in Sec. II B, the shape features nodes at the center of the bay. (d) An arbitrary attenuated wave in the lower region of the first stop band. (e) The lower edge of the second passband. The beads are at rest and string segments are vibrating at their own natural frequencies. This is a mode of finite string systems with uniform bay length. (f) An arbitrary second passband waveform. The beads are featuring the same global shape as in (b) but the string segment connecting adjacent beads is now exhibiting vibration in its first mode. (g) A string vibrating at the frequency of the upper edge of the second passband. The shape features zero slope at the center of the bay as predicted. (h) An arbitrary attenuated wave in the second stop band. The frequency is close to the upper limit of the stop band hence each individual string segment has begun undergoing second mode vibration.

Figure 5 depicts, at various frequencies, a right traveling wave with amplitude 1.0 incident to the left end of an infinite string with coupling value $\bar{R}=1.0$. Given the dimensionless frequency $\bar{\omega}$ and the dimensionless coupling \bar{R} , then the wave number k and exponential decay γ may be determined and the complex amplitude of all beads due to an incoming wave may be calculated. The deflection of the internal points of the string may be recovered using Eq. (14).

C. Modes of the finite ordered system

In the previous section an infinite periodic beaded string was examined. We now turn our attention to a finite system, fixed at both ends.

The natural frequencies of a finite ordered system can be found in an elegant manner using transfer matrices. The dynamics of a string with N ordered bays with $N-1$ beads and fixed ends are represented by:

$$\begin{aligned} \begin{bmatrix} \bar{y}_N \\ \bar{y}_{N-1} \end{bmatrix} &= \mathbf{Z}_0^{N-1} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_0 \end{bmatrix} \\ &= \mathbf{X} \mathbf{W}^{N-1} \mathbf{X}^{-1} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_0 \end{bmatrix} \\ &= \frac{\lambda}{\lambda^2 - 1} \begin{bmatrix} 1 & 1 \\ \lambda^{-1} & \lambda \end{bmatrix} \begin{bmatrix} \lambda^{N-1} & 0 \\ 0 & \lambda^{1-N} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \lambda & -1 \\ -\lambda^{-1} & 1 \end{bmatrix} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_0 \end{bmatrix}. \end{aligned} \quad (31)$$

This relationship may be rearranged to the form

$$\begin{bmatrix} \bar{y}_N \\ \bar{y}_0 \end{bmatrix} = \mathbf{B} \begin{bmatrix} \bar{y}_{N-1} \\ \bar{y}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (32)$$

which yields a nontrivial solution if $\det[\mathbf{B}]=0$. This simplifies to, after some algebra,

$$\frac{\lambda^{2N} - 1}{\lambda^{2(N-2)} - 1} = 0, \quad (33)$$

or

$$\lambda = 2^{N\sqrt{1}} \quad \text{and} \quad \lambda \neq 2^{(N-2)\sqrt{1}}, \quad (34)$$

which when combined gives

$$k = \frac{i\pi}{N}, \quad \text{where} \quad i = 1, \dots, N-1 \quad (35)$$

and $\gamma=0$, placing the natural frequencies squarely in the passbands. The frequency equation takes the form

$$2 \cos \bar{\omega} - \frac{\bar{\omega}}{R} \sin \bar{\omega} = 2 \cos \frac{i\pi}{N}, \quad i = 1, \dots, N-1. \quad (36)$$

It does not admit the edge frequencies $\omega=n\pi$ as solutions although, as we observed in Sec. I A, $\omega=n\pi$ is indeed a natural frequency. The reason that it is not accounted for is that the transfer matrix that was used is singular for this value of the frequency.

It is of interest to compare the results outlined here with the work of Mead.² Mead's classification scheme for periodic systems is based on symmetry and fixed points in the

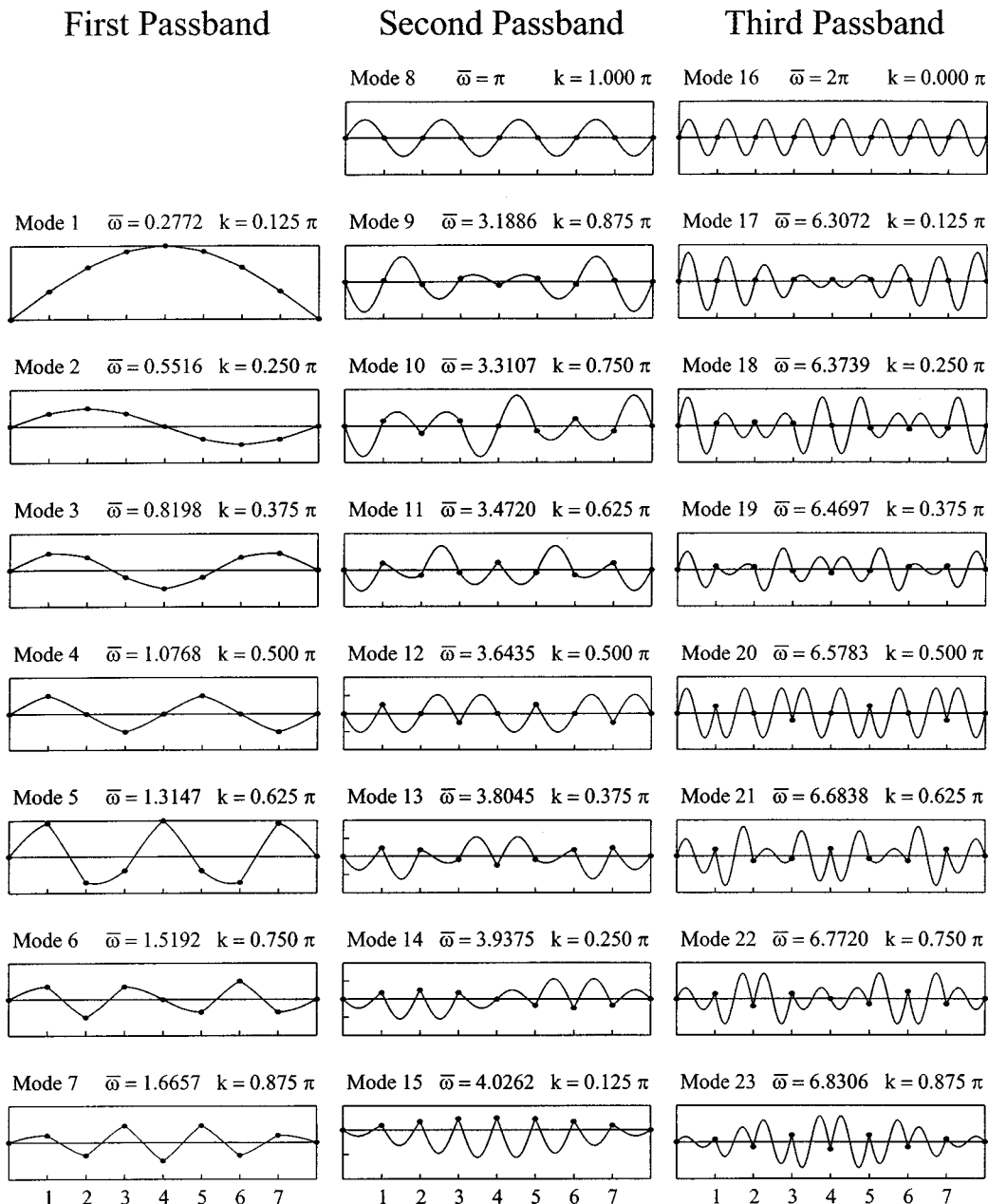


FIG. 6. The mode shapes in the first three passbands of an ordered seven-bead system with $\bar{R} = 1.0$.

subelements of the system. The beaded string is symmetric because even though the elements used here are unsymmetric, the system might just as well have been modeled with symmetric elements, e.g., one with half a bead at each end of a string element. The subelements of the beaded string have no fixed points. Mead predicts that symmetric periodic systems without fixed point will have $(N-1)$ natural frequencies within each of the passbands and one additional natural frequency at a passband edge. This agrees exactly with our findings.

The natural frequencies of any finite, ordered system may be calculated from the frequency equation (36) and the corresponding mode shape may be generated using the dispersion relation [Eq. (25)]. Hodges and Woodhouse¹⁰ examined a seven-bead system of unspecified coupling, whose

mode shapes in the first two passbands compare perfectly with the results illustrated in Fig. 6.

III. THE DISORDERED SYSTEM

The presence of disorder can drastically alter the system's behavior through a phenomenon called localization. The following sections present the effect of disorder on the modal structure of the loaded string. The mode localization is quantified by studying propagation of waves in infinite disordered systems.

A. Modes of the disordered system

In Sec. II B the natural frequencies and modes of an ordered system were studied. Natural frequencies were found

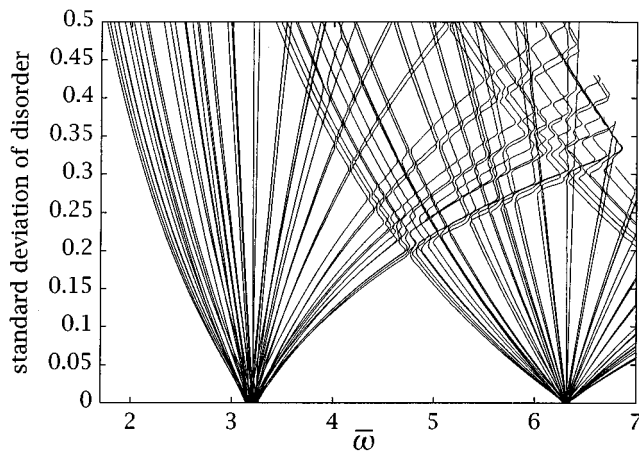


FIG. 7. Spreading and veering of the second and third natural frequency clusters as the strength of bead-spacing disorder is varied. The system has 50 bays with a given disorder pattern and $\bar{R}=0.01$.

to be located in the passbands with extended (i.e., *spatially periodic*) mode shapes. We now wish to examine the effect of disorder on the modes of the system. Solving Eq. (17) for nontrivial solutions yields the natural frequencies of a finite system with any given disorder.

Figure 7 depicts the variation of the natural frequencies in the second and third passband clusters of a 50 bay system with $\bar{R}=0.01$ as the strength of the bead-spacing disorder is varied. The study is based on a given pattern of random numbers, only the standard deviation of the disorder is var-

ied. Initially one notices the fanning out of natural frequencies as the disorder strength is increased. At $\sigma_\delta \approx 0.2$ the second and third passband have merged. The fanning out of natural frequencies is to be expected since the shorter (stiffer) bays now lead to lower frequencies, whereas the longer than average bays lead to higher frequencies. Thus, some of the modes have natural frequencies in the frequency range that corresponds to the stop band of the ordered structure. Upon closer inspection of Fig. 7 some interesting features are revealed, namely the multiple veerings of curves corresponding to different eigenvalues, first within each cluster and, for higher levels of disorder, due to cluster interaction. For a nice discussion of the mechanism behind the curve veering see Ref. 15.

Figure 8 illustrates localized modes of an eight-bay system with two levels of coupling. Mode localization is apparent even when the bead-spacing disorder is only 1% ($\sigma_\delta = 0.01$). Note how localization effects are more pronounced in mode 1 and 7 than in modes 3 and 5, since the former are closer to passband edges.

B. Wave localization

The effects of disorder are best understood by comparing the propagation of incident waves in ordered and disordered systems. In Sec. II A it was demonstrated how the matrix of wave modes diagonalizes the transfer matrix and hence that right and left traveling waves are independent. In

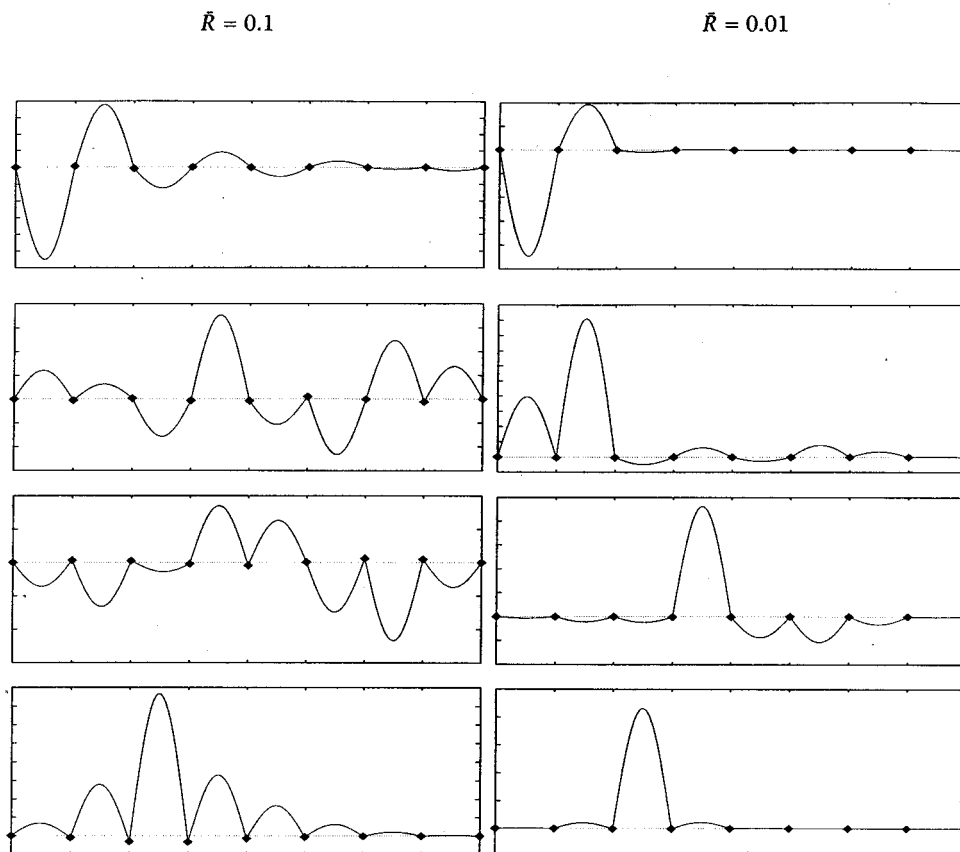


FIG. 8. Modes 1 (top), 3, 5, and 7 of the second mode cluster of two, weakly coupled, seven-bead loaded string with 1% bead-spacing disorder.

the disordered case, the matrix \mathbf{X} does not, in general, diagonalize the transfer matrix, $\mathbf{Z}_{i,i-1}$. Instead we have

$$\begin{bmatrix} L \\ R \end{bmatrix}_i = \mathbf{W}_i \begin{bmatrix} L \\ R \end{bmatrix}_{i-1} = \begin{bmatrix} \frac{1}{t_i} & -\frac{r_i}{t_i} \\ \frac{r_i}{t_i} & t_i - \frac{r_i^2}{t_i} \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix}_{i-1}, \quad (37)$$

where t_i is the transmission coefficient for the i th bay interface and r_i is the reflection coefficient. An ordered assembly is such that $\gamma_i = 0$, thus waves travel unreflected. In the disordered assembly, there is reflection, or scattering, at the interfaces between the dissimilar bays. The off-diagonal elements in \mathbf{W}_i give the portion of the wave that is reflected. A wave incident to a segment of randomly disordered bays will experience multiple reflections whose effect may be to trap the wave, such that only a fraction of an incident wave is transmitted along to the far end of the disordered segment. This effect is called localization. In a disordered system passbands no longer exist since all waves are attenuated.

For a segment of N dissimilar bays, the wave transfer matrix is the product of the random wave transfer matrices of the individual bays

$$\mathcal{W}_N = \prod_{i=1}^N \mathbf{W}_i = \begin{bmatrix} \frac{1}{\tau_N} & -\frac{\rho_N}{\tau_N} \\ \frac{\rho_N}{\tau_N} & \tau_N - \frac{\rho_N^2}{\tau_N} \end{bmatrix}. \quad (38)$$

The one-one term in the product, $1/\tau_N$, indicates which portion of an incident wave is transmitted to the far end of the disordered segment. Obviously the behavior of the transmission coefficient for the disordered segment τ_N governs the strength of the effects of disorder and the resulting localization. We define the *localization factor* as

$$\gamma = \lim_{N \rightarrow \infty} \left[\frac{1}{N} \ln \left| \frac{1}{\tau_N} \right| \right], \quad (39)$$

implying that asymptotically, the ratio of emergent to incident wave decreases exponentially with an increasing number of bays, N . The localization factor γ defines the average exponential decay rate per bay and thus is a descriptor of the strength of localization. Assuming that the random variables δ_i and ϵ_i ($i=1, \dots, N$) from Eq. (7) form an ergodic sequence, then the transmission coefficient τ_N which is a function of $\delta_1, \dots, \delta_N$, is also ergodic. Ergodicity implies that, with probability 1, the limit in Eq. (39) is equivalent to the ensemble average for a finite disordered segment of length N :

$$\gamma = \left\langle \frac{1}{2N} \ln \left| \frac{1}{\tau_N} \right|^2 \right\rangle. \quad (40)$$

where $\langle \rangle$ denotes the expected value of a random variable.

The evaluation of a localization factor may also be accomplished via an alternate vibration formulation, as outlined in Sec. I B, by examining the amplitude of one end of the string when the other end is driven. This method was originally proposed by Herbert and Jones¹⁶ and later used by Hodges and Woodhouse.¹⁰ This involves calculating the average exponential decay per bay in a disordered system, as

$$\gamma_N = -\frac{1}{N} \ln \left| \frac{y_1}{Y} \right|. \quad (41)$$

where Y is the prescribed amplitude of the driven end. Finally, averaging yields

$$\gamma = \langle \gamma_N \rangle = \lim_{N \rightarrow \infty} \gamma_N. \quad (42)$$

In general the expressions for the localization factor cannot be evaluated in closed form. We thus seek analytical approximations using perturbations. The two sections that follow outline two perturbation techniques available for this purpose. In Sec. III C the disordered system is regarded as a perturbation of the ordered system by expanding in terms of the small disorder parameter. Section III D provides an alternative method in the cases where the expansion in Sec. III C is nonuniform, by regarding the weakly coupled disordered system as a perturbation of an uncoupled system.

C. Weak localization

When the ratio of disorder to coupling is small, the disordered system is treated as a perturbation of an ordered system. A transfer matrix formulation is used to study the propagation of waves in a disordered beaded string with an infinite number of bays. A transfer matrix for a disordered bay is expanded in the small disorder parameters, δ and ϵ . Using the displacement/slope representation of Eq. (1), we have

$$\mathbf{T}_i = \begin{bmatrix} \cos \bar{\omega}_i & \frac{\sin \bar{\omega}_i}{\bar{\omega}} \\ -\bar{\omega} \left(\sin \bar{\omega}_i + \frac{\bar{\omega}_i}{\bar{R}_i} \cos \bar{\omega}_i \right) & \cos \bar{\omega}_i - \frac{\bar{\omega}_i}{\bar{R}_i} \sin \bar{\omega}_i \end{bmatrix}, \quad (43)$$

where

$$\bar{\omega}_i = \bar{\omega}(1 + \delta_i) \quad \text{and} \quad \bar{R}_i = \bar{R} \frac{(1 + \delta_i)}{(1 + \epsilon_i)}. \quad (44)$$

Noting that \mathbf{T}_i is linear in ϵ_i , a second-order expansion with respect to the disorder yields

$$\mathbf{T}_i = \mathbf{T}_o + \delta_i \mathbf{T}_\delta + \frac{\delta_i^2}{2} \mathbf{T}_{\delta^2} + \epsilon_i \mathbf{T}_\epsilon + O(\delta_i^3), \quad (45)$$

where \mathbf{T}_o is the transfer matrix of the ordered structure and

$$\mathbf{T}_\delta = \left. \frac{\partial}{\partial \delta_i} \mathbf{T}_i \right|_{\delta_i=0, \epsilon_i=0} = \begin{bmatrix} -\bar{\omega} \sin \bar{\omega} & \cos \bar{\omega} \\ \frac{\bar{\omega}^3}{\bar{R}} \sin \bar{\omega} - \bar{\omega}^2 \cos \bar{\omega} & -\frac{\bar{\omega}^2}{\bar{R}} \cos \bar{\omega} - \bar{\omega} \sin \bar{\omega} \end{bmatrix},$$

$$\mathbf{T}_{\delta^2} = \frac{\partial^2}{\partial \delta_i^2} \mathbf{T}_i \Big|_{\substack{\delta_i=0 \\ \epsilon_i=0}} = \begin{bmatrix} -\frac{\bar{\omega}^2}{R} \cos \bar{\omega} & -\frac{\bar{\omega}}{R} \sin \bar{\omega} \\ \frac{\bar{\omega}^4}{R} \cos \bar{\omega} + \bar{\omega}^3 \sin \bar{\omega} & \frac{\bar{\omega}^3}{R} \sin \bar{\omega} - \bar{\omega}^2 \cos \bar{\omega} \end{bmatrix}, \quad (46)$$

$$\mathbf{T}_{\epsilon} = \frac{\partial}{\partial \epsilon_i} \mathbf{T}_i \Big|_{\substack{\delta_i=0 \\ \epsilon_i=0}} = \begin{bmatrix} 0 & 0 \\ -\frac{\bar{\omega}^2}{R} \cos \bar{\omega} & -\frac{\bar{\omega}}{R} \sin \bar{\omega} \end{bmatrix}$$

are Taylor coefficient matrices. Each of the above matrices may be transformed into wave coordinates using the transformation

$$\mathbf{W} = \mathbf{X}^{-1} \mathbf{S}_o \mathbf{T} \mathbf{S}_o^{-1} \mathbf{X}, \quad (47)$$

where the matrix \mathbf{X} is the matrix of wave modes, defined in Eq. (22), and \mathbf{S}_o is the (ordered system) transformation from $[\bar{y}, \phi^R]_i$ coordinates to $[\bar{y}_i, \bar{y}_{i-1}]^T$ coordinates defined in Eq. (10). The matrix \mathbf{T} in Eq. (47) refers to each of the matrices in Eq. (46).

As outlined in the previous section, the wave transfer matrix for an assembly of N disordered bays must be formulated. The system wave transfer matrix \mathcal{W}_N is, from Eq. (38)

$$\mathcal{W}_N = \prod_{i=1}^N \left(\mathbf{W}_o + \delta_i \mathbf{W}_{\delta} + \epsilon_i \mathbf{W}_{\epsilon} + \frac{\delta_i^2}{2} \mathbf{W}_{\delta^2} + O(\delta_i^3) \right). \quad (48)$$

Carrying out the multiplication to the second order yields

$$\begin{aligned} \mathcal{W}_N &= \mathbf{W}_o^N + \sum_{i=1}^N \mathbf{W}_o^{i-1} \mathbf{W}_{\delta} \mathbf{W}_o^{N-i} \delta_i \\ &+ \sum_{i=1}^N \mathbf{W}_o^{i-1} \mathbf{W}_{\epsilon} \mathbf{W}_o^{N-i} \epsilon_i + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{W}_{ij}^{\epsilon \epsilon} \epsilon_i \epsilon_j \\ &+ \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{W}_{ij}^{\delta \epsilon} \delta_i \epsilon_j + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{W}_{ij}^{\delta \delta} \delta_i \delta_j \\ &+ \sum_{i=1}^N \mathbf{W}_o^{i-1} \mathbf{W}_{\delta^2} \mathbf{W}_o^{N-i} \frac{\delta_i^2}{2} + O(\delta_i^3). \end{aligned} \quad (49)$$

The matrices $\mathbf{W}_{ij}^{\delta \delta}$, $\mathbf{W}_{ij}^{\delta \epsilon}$, and $\mathbf{W}_{ij}^{\epsilon \epsilon}$ have a complex form that need not be evaluated since those terms will vanish in the averaging process that follows due to the independence of the random variables δ_i and ϵ_i ($i=1, \dots, N$). For an approximation of the localization factor, only the first diagonal element of $\mathcal{W}_N, 1/\tau_N$, is required. It is found to be

$$\frac{1}{\tau_N} = e^{jN} \left[1 + a j \sum_{i=1}^N \delta_i + b j \sum_{i=1}^N \epsilon_i + c \sum_{i=1}^N \frac{\delta_i^2}{2} \right], \quad (50)$$

$a, b, c \in \mathbb{R},$

where a , b , and c are real valued expressions derived by the algebra:

$$a^2 = \bar{\omega}^2 + \frac{\bar{\omega}^4}{4\bar{R}^2 \sin^2 k}, \quad b^2 = \left(\frac{\bar{\omega}}{2\bar{R}} \frac{\sin \bar{\omega}}{\sin k} \right)^2, \quad (51)$$

$$c = -\bar{\omega}^2.$$

The fact that the terms in Eq. (50) are either real or pure imaginary greatly simplifies the evaluation of the modulus of $1/\tau_N$. Assuming $\ln|1+x| \approx x$, for small x , then Eq. (40) yields

$$\begin{aligned} \gamma &= \left\langle \frac{1}{2N} \ln \left| \frac{1}{\tau_N} \right|^2 \right\rangle \\ &\approx \frac{a^2}{2} \left\langle \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \delta_i \delta_j \right\rangle \\ &+ \frac{b^2}{2} \left\langle \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \epsilon_i \epsilon_j \right\rangle + c \left\langle \frac{1}{N} \sum_{i=1}^N \frac{\delta_i^2}{2} \right\rangle, \end{aligned} \quad (52)$$

where $\langle \rangle$ denotes an average. After some algebra Eq. (52) reduces to

$$\begin{aligned} \gamma &= \frac{\bar{\omega}^2}{8\bar{R}^2 \sin^2 k} (\bar{\omega}^2 \sigma_{\delta}^2 + \sigma_{\epsilon}^2 \sin^2 \bar{\omega}) \\ &= \frac{\bar{\omega}^2 (\bar{\omega}^2 \sigma_{\delta}^2 + \sigma_{\epsilon}^2 \sin^2 \bar{\omega})}{2\bar{R}^2 (2 - \beta_o)(2 + \beta_o)}, \end{aligned} \quad (53)$$

where, as before, $\beta_o = 2 \cos \bar{\omega} - (\bar{\omega}/R) \sin \bar{\omega}$. In the middle of the passband, where $k = \pi/2$, we have $\beta_o = 0$ and hence,

$$2 \cos \bar{\omega}_{\text{mid}} = \frac{\bar{\omega}_{\text{mid}}}{R} \sin \bar{\omega}_{\text{mid}} \quad (54)$$

and Eq. (53) simplifies to

$$\gamma_{\text{mid}} = \frac{\bar{\omega}_{\text{mid}}^4}{8\bar{R}^2} \sigma_{\delta}^2 + \frac{\sigma_{\epsilon}^2}{2} \cos^2 \bar{\omega}_{\text{mid}} \quad \left(k = \frac{\pi}{2} \right). \quad (55)$$

Focusing first on the effect of the disorder of the bead mass ($\sigma_{\delta} = 0$; $\sigma_{\epsilon} \neq 0$) reveals interesting features. The mid-passband localization factor, $\gamma_{\text{mid}}^{\epsilon}$, depends on \bar{R} through Eq. (54). As \bar{R} decreases the lower passband edge approaches the upper edge, $\bar{\omega} = n\pi$. This causes saturation as $\cos^2 \bar{\omega}_{\text{mid}} \rightarrow 1$, i.e.,

$$\gamma_{\text{mid}}^{\epsilon} \rightarrow \frac{\sigma_{\epsilon}^2}{2} \quad \text{when} \quad \bar{R} \rightarrow 0 \quad \text{at} \quad k = \frac{\pi}{2}. \quad (56)$$

This suggests that localization due to bead-mass disorder is always weak and that the classical perturbation assumptions are never violated. Another interesting feature of bead-mass disorder is the lack of localization at the lower edge of what was the passband for the system's ordered counterpart. As was previously explained, these are the natural frequencies of a string with fixed ends and thus correspond to the beads being at rest. Obviously, disordering the bead mass has no effect at these frequencies. Equation (53) confirms this, since

$$\gamma^{\epsilon} = \frac{\sigma_{\epsilon}^2 \bar{\omega}^2 \sin \bar{\omega}}{(8\bar{R}^2 - 2\bar{\omega}^2) \sin \bar{\omega} + 8\bar{\omega}\bar{R} \cos \bar{\omega}} \rightarrow 0 \quad \text{as} \quad \bar{\omega} \rightarrow n\pi. \quad (57)$$

At the upper passband edge the approximate localization factor goes to infinity and Eq. (53) loses its validity, due to the double eigenvalue of the \mathbf{T}_o matrix and the resulting singularity of the \mathbf{X} matrix at the passband edges.

Turning our attention to the effect of bead-spacing disorder ($\sigma_\delta \neq 0$; $\sigma_\epsilon = 0$) we find, by Eq. (53), that the localization factor becomes unbounded at passband edges, where $\beta_o = \pm 2$. Also apparent, from Eq. (55), is the fact that as interbay coupling \bar{R} is decreased, the localization factor increases, possibly violating the assumptions associated with perturbing away from the ordered system. The case of the first passband, however, is different and is treated below. Using Eq. (54), Eq. (55) simplifies to

$$\gamma_{\text{mid}} = \frac{\bar{\omega}_{\text{mid}}^2}{2 \tan^2 \bar{\omega}_{\text{mid}}} \sigma_\delta^2. \quad (58)$$

Using the same reasoning as in the bead-mass disorder case, we find that localization factor at the first midband goes to $\sigma_\delta^2/2$ as $\bar{R} \rightarrow 0$. This, in addition to the fact that the localization due to bead-spacing disorder is zero at $\bar{\omega} = 0$, indicates that bead-spacing disorder effects in the first passband are identical to bead-mass disorder. In the case of the bead-mass disorder it was argued that for weak coupling the beads do not participate in the motion and disorder has no effect. Similarly in the first passband of a weakly coupled system the string segments do not participate in the motion [see Fig. 5(a)], so spacing disorder has no effect.

It is tempting to extrapolate these findings by hypothesizing that whenever a periodic system is immune to parameter disorder at one or more frequencies of a passband, it will only experience weak localization in this passband. This is true here for localization due to bead-mass disorder in all passbands and for localization due to bead-spacing disorder in the first passband. This hypothesis is further supported by the experience from another family of periodic systems, e.g., a series of oscillators connected by springs (see Ref. 17). In this type of system localization due to disorder in the stiffness of a spring coupling the oscillators only causes weak localization. At one of the frequencies in each passband the oscillators vibrate in phase, thereby not extending the coupling springs. This mode of vibration is independent of the coupling spring stiffness and wave localization is zero. If correct, this hypothesis could provide intuition regarding sensitivity to parameter mistuning.

D. Strong localization (spacing disorder only)

If the localization factor as predicted by the classical perturbation analysis becomes large, as is the case for small values of \bar{R} in Eq. (55), the foundation of the classical perturbation method collapses. The previous section discussed how this will not occur in the bead-mass disorder case.

When bead spacing is disordered and interbay coupling is weak, a different technique must be used to approximate the localization factor. A vibration approach is used, studying the amplitude at one end of the string when the other end is driven. From Sec. II B the matrix equation of motion for a N degree of freedom string with one end fixed and the other driven by $y_o = Y e^{j\Omega t}$ is,

$$\text{tridiag}[1; -\beta_{i,i-1}; \alpha_{i,i-1}] \mathbf{y} = [\mathbf{A}] \mathbf{y} = \mathbf{f}. \quad (59)$$

From Cramer's rule, the maximum relative displacement of the bead adjacent to the fixed end is

$$\frac{y_1}{Y} = \mathbf{A}_{(N,1)}^{-1} = \frac{[\text{cofactor matrix}]_{(1,N)}}{\det[\mathbf{A}]}, \quad (60)$$

as long as the driving frequency Ω is not a natural frequency of the system. The appropriate minor determinant of \mathbf{A} is simply $(-1)^{N-1}$, and the determinant of \mathbf{A} may be written as the product of the eigenvalues of \mathbf{A} , λ_i , thus

$$\left| \frac{y_1}{Y} \right| = |\mathbf{A}_{(N,1)}^{-1}| = \left(\prod_{i=1}^N \lambda_i \right)^{-1}. \quad (61)$$

Using the fact that when interbay coupling is weak, terms containing \bar{R} in the denominator are larger than terms that do not, a modified perturbation method is used to calculate the eigenvalues of \mathbf{A} . Treating the coupling parameter \bar{R} as a small disorder parameter is tantamount to treating the weakly coupled system as a perturbation of a series of uncoupled string segments. Then

$$\mathbf{A} = \mathbf{A}_o + \delta \mathbf{A}, \quad (62)$$

where, given that $\epsilon_{i-1} = 0$,

$$\mathbf{A}_o = \text{diag} \left[\frac{\bar{\omega}}{\bar{R}} \sin \bar{\omega}_i \right] \quad (63)$$

and

$$\delta \mathbf{A} = \text{tridiag} \left[1; -\cos \bar{\omega}_i - \cos \bar{\omega}_{i-1} \frac{\sin \bar{\omega}_i}{\sin \bar{\omega}_{i-1}}; \frac{\sin \bar{\omega}_i}{\sin \bar{\omega}_{i-1}} \right]. \quad (64)$$

Since $\delta \mathbf{A}$ is small compared to \mathbf{A}_o , it can be shown that the eigenvalues of \mathbf{A} may be approximated as the sum of the eigenvalues of \mathbf{A}_o and the diagonal values of $\delta \mathbf{A}$,

$$\lambda_i = \frac{\bar{\omega}}{\bar{R}} \sin \bar{\omega}_i - \cos \bar{\omega}_i - \cos \bar{\omega}_{i-1} \frac{\sin \bar{\omega}_i}{\sin \bar{\omega}_{i-1}}, \quad (65)$$

and then

$$\gamma_N = -\frac{1}{N} \ln \left| \frac{y_1}{Y} \right| = \frac{1}{N} \ln \left| \prod_{i=1}^N \lambda_i \right| = \frac{1}{N} \sum_{i=1}^N \ln |\lambda_i|. \quad (66)$$

The expected value of γ is, assuming for simplicity, that the random variable δ is uniformly distributed over $[-W, W]$

$$\gamma = \frac{1}{4W^2} \int_{-W}^W \int_{-W}^W \ln |\lambda_i| d\delta_{i-1} d\delta_i. \quad (67)$$

This equation does not allow a closed-form solution and must be solved numerically. However, since it has already been assumed that \bar{R} is small, one wonders if the approximation

$$\lambda_i \approx \frac{\bar{\omega}}{\bar{R}} \sin \bar{\omega}_i \quad (68)$$

might not be appropriate. This amounts to ignoring $\delta\mathbf{A}$ above, a zeroth-order expansion of \mathbf{A} . This approximation yields

$$\gamma \approx \frac{1}{4W} \int_{-W}^W \ln \left| \frac{\bar{\omega}}{R} \sin[\bar{\omega}(1+\delta)] \right| d\delta, \quad (69)$$

which also eludes a closed-form solution but may be numerically integrated at substantial savings.

E. Results and Monte Carlo simulation

Verification of the approximate theoretical values obtained for the localization factor was accomplished by Monte Carlo simulations. Simulations were performed using a transfer matrix approach except near the passband edges where the \mathbf{X} matrix becomes singular. For those frequencies a vibration formulation was used. Simulation by transfer matrices of a string with N bays required the generation of a series of random transfer matrices \mathbf{Z}_i and successive matrix multiplications. The resulting product \mathcal{Z}_N has transformed into a complex wave transfer matrix, \mathcal{W}_N and a localization factor calculated from

$$\gamma_N = \frac{-\ln|\omega'_{1,1}|}{N} \quad (70)$$

and subsequently averaging γ_N for a large number of realizations. In order to reduce numerical errors caused by the matrix multiplication, a system smaller than 50 bays was used and as many as 10 000 realizations were needed to achieve satisfactory convergence of the average of the localization factor. Simulation by a vibration approach, on the other hand, involves solving Eq. (16) for y_1 . The entries of the system matrix are again functions of the random variables and are generated with random number generators. Systems with 100 bays were analyzed to lessen end effects. Adequate convergence was reached in fewer than 5000 realizations for the frequencies involved. The localization factor for each realization was computed by

$$\gamma_N = -\frac{1}{N} \ln \left| \frac{y_1}{y_o} \right| \quad (71)$$

and subsequently averaging for multiple realizations.

The results of the simulations are illustrated in Figs. 9–12. Figures 9 and 10 show the effect of coupling. The localization factor at the passband center, γ_{mid} , is plotted as a function of the dimensionless coupling \bar{R} . Figure 9 offers confirmation that bead-mass disorder leads to weak localization only and that saturation occurs. A probable justification of the lack of strong localization in the case of bead-mass disorder is that as the coupling vanishes, the beads cease to participate in the motion of the string so that a variation in the mass of the beads has a lessened effect. For spacing disorder, Fig. 10 shows the excellent agreement between the simulated values and the theoretical values in the limiting cases of strong and weak coupling. Also shown is a curve generated by the modified formula of Eq. (69), which consistently showed agreement with the simulated values well beyond what was expected.

Figures 11 and 12 feature the localization factor in a weakly coupled system ($\bar{R}=0.1$) for bead-mass disorder and

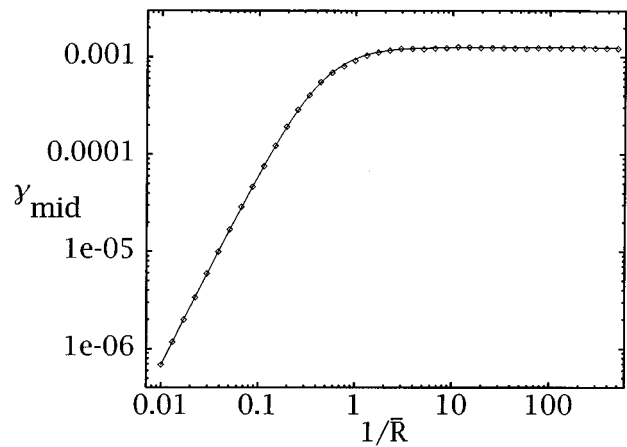


FIG. 9. Effect of bead-mass disorder ($\sigma_\epsilon=0.05$). The classical perturbation approximation of the second midband localization factor (line) and the Monte Carlo simulation results (dots) are plotted as a function of inverse coupling.

bead spacing disorder, respectively, plotted as a function of frequency. The figures compare the statistical perturbation approximations of the localization factor and Monte Carlo simulations. The exponential attenuation factor for the ordered system has also been included for comparison. The attenuation in the ordered system's stop bands clearly dominates any localization effects, as, in these frequency regions, the Monte Carlo simulations of the disordered system agree closely with the exponential attenuation rate. Of course, the ordered and disordered systems' results are vastly different in the ordered system's passbands. Waves in the disordered system are attenuated at all frequencies. Figure 11 illustrates how in the case of bead-mass disorder the localization factor is zero at the lower passband edge and reaches only low values throughout the passband, even for relatively high values of disorder. The classical perturbation gives an accurate approximation of the localization factor throughout the passband, as predicted by Eq. (55).

Figure 12 depicts the much richer effects of spacing disorder. In the first passband, the loaded string exhibits global

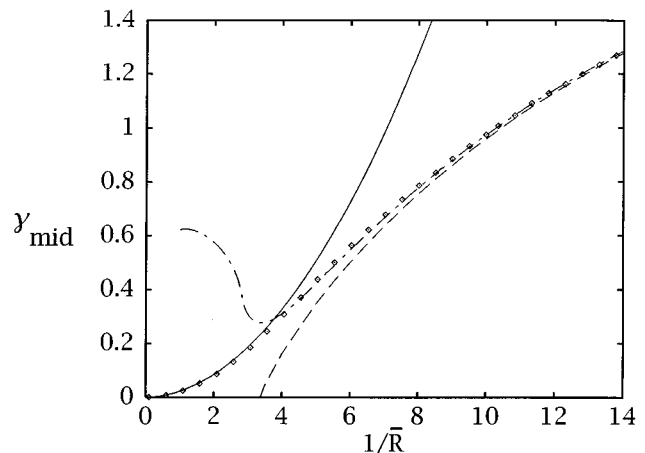


FIG. 10. Effect of spacing disorder ($\sigma_\delta=0.05$). The classical perturbation approximation of the second midband localization factor (—), modified perturbation approximation (---), approximate modified approach [Eq. (69)] (-.-), and the Monte Carlo simulation results (dots) are plotted as a function of inverse coupling.

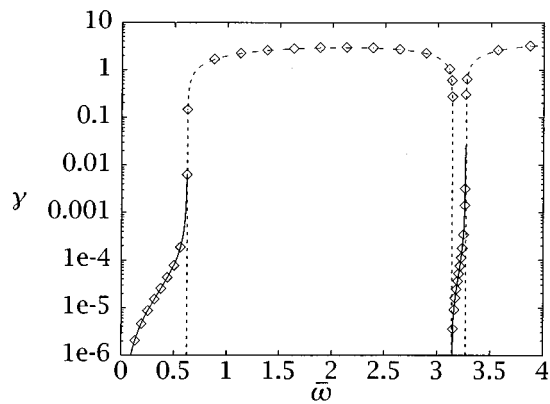


FIG. 11. Localization factor due to 5% bead-mass disorder plotted as a function of dimensionless frequency. The classical perturbation approximation of the localization factor (—) and the Monte Carlo simulation (dots) are shown alongside the ordered system exponential decay value in the stop band (---). $\bar{R}=0.1$.

motion, without participation of the individual string segments. Here the string features only weak localization much like that evidenced for the bead-mass disorder case. The classical perturbation method approximates accurately the localization factor throughout the passband. This is also the case in the second passband, although much higher levels of localization are observed. In the third passband the localization factor has reached such high values that the classical perturbation approach is no longer valid and overestimates the localization factor. Here the modified approach is accurately approximating the localization factor.

He and Maynard¹¹ arrived at some of the same conclusions in their experimental investigation of the beaded string. The experimental setup featured a vertically suspended copper wire with attached masses, capable of being excited longitudinally or transversely. The effects of bead-mass disorder (which He and Maynard called alloy-type disorder) and bead-spacing disorder (which they called liquid-type disorder) were both studied. The effect of bead-spacing disorder was found to be “dramatic” compared to “negligible” ef-

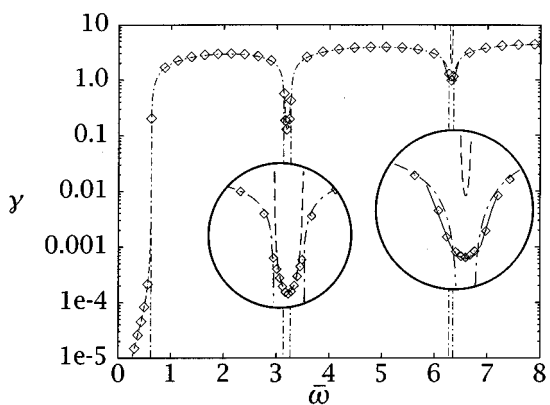


FIG. 12. Localization factor due to 1% spacing disorder plotted as a function of frequency. The classical perturbation approximation of the localization factor (---), the modified perturbation approximation (— · —), and the Monte Carlo simulation (dots) are plotted alongside the ordered system exponential decay value in the stop band (— — —). Inserts show detailed view of passbands. $\bar{R}=0.1$.

fects of bead-mass disorder, a good description of strong and weak localization.

IV. CONCLUSIONS

Numerical simulations confirm analytical approximations of the localization factor in a string with attached beads. The effects of disorder on both the bead-mass and the bead-spacing was examined. It has been established that the loaded string exhibits only weak localization for bead-mass disorder and that the localization factor saturates as coupling is decreased to zero. The classical perturbation approximation of the localization factor is accurate for all values of interbay coupling for bead-mass disorder. For spacing disorder, however, both weak and strong localization are evidenced, depending on coupling and disorder strength. This result has been observed experimentally by He and Maynard.¹¹ For spacing disorder two localization factor approximations are valid in the limiting cases of strong and weak coupling. In all cases the results compared favorably with Monte Carlo simulations. The effect of disorder on the natural frequencies and mode shapes of the loaded string was also presented. The natural frequencies that lie inside the passbands of the ordered system were found to fan out as disorder was introduced. Mode localization was apparent.

- ¹I. B. Crandall, *Theory of Vibrating Systems and Sound* (Van Nostrand, New York, 1926).
- ²D. J. Mead, “Wave propagation and natural modes in periodic systems, I: mono-coupled systems,” *J. Sound Vib.* **40**, 1–18 (1975).
- ³N. A. Valero and O. O. Bendiksen, “Vibration Characteristics of Mistuned Shrouded Blade Assemblies,” *ASME J. Eng. Gas Turbines Power* **108**, 293–299 (1986).
- ⁴S. T. Wei and C. Pierre, “Localization Phenomena in Mistuned Assemblies with Cyclic Symmetry, Part I: Free Vibrations,” *ASME J. Vib. Acoust., Stress Reliability Design* **110**, 429–438 (1988).
- ⁵D. Bouzit and C. Pierre, “Vibration Confinement Phenomena in Disordered, Mono-coupled, Multi-span Beams,” *ASME J. Vib. Acoust.* **114**, 521–530 (1992).
- ⁶C. Pierre and P. D. Cha, “Strong mode localization in nearly periodic disordered structures,” *AIAA J.* **27**, 227–241 (1989).
- ⁷C. Pierre and E. H. Dowell, “Localization of vibrations by structural irregularity,” *J. Sound Vib.* **114**, 529–564 (1987).
- ⁸P. J. Cornwell and O. O. Bendiksen, “Localization of vibrations in large space reflectors,” *AIAA J.* **27**, 219–226 (1987).
- ⁹G. J. Kissel, “Localization in Disordered Periodic Structures,” Ph.D. thesis, Massachusetts Institute of Technology, 1988.
- ¹⁰C. H. Hodges and J. Woodhouse, “Vibration isolation from irregularity in a nearly periodic structure: theory and measurements,” *J. Acoust. Soc. Am.* **74**, 894–905 (1983).
- ¹¹S. He and J. D. Maynard, “Detailed Measurements of Inelastic Scattering in Anderson Localization,” *Phys. Rev. Lett.* **57**, 3171–3174 (1986).
- ¹²G. Rajagopal, “Mode Localization in Almost Periodic Structures: Theory and Experiments for the Case of Masses on a String,” in *Proceedings of the Second International Offshore and Polar Engineering Conference*, San Francisco, California, 14–19 June 1992 (International Society of Offshore and Polar Engineers), Vol. II, pp. 426–433.
- ¹³S. S. Sergeev and W. D. Iwan, “The Natural Frequencies and Mode Shapes of Cables with Attached Masses,” *ASME J. Energy Resources Technol.* **103**, 237–242 (1981).
- ¹⁴E. H. Dowell, “Free vibrations of an arbitrary structure in terms of component modes,” *ASME J. Appl. Phys.* **4**, 1145–1161 (1972).
- ¹⁵C. Pierre, “Mode Localization and Eigenvalue Loci Veering Phenomena in Disordered Structures,” *J. Sound Vib.* **126**, 485–502 (1988).
- ¹⁶D. C. Herbert and R. R. Jones, “Localized States in Disordered Systems,” *J. Phys. C: Solid State Phys.* **4**, 1145–1161 (1971).
- ¹⁷G. S. Ottarsson and C. Pierre, “A Transfer Matrix Approach to Vibration Localization in Mistuned Blade Assemblies,” in *Proceedings of the International Gas Turbine and Aeroengine Congress*, number 93-GT-115, Cincinnati, Ohio, 1993.

Diffuse transmission of structure-borne sound at periodic junctions of semi-infinite plates

I. Bosmans and G. Vermeir

Laboratory of Building Physics, Department of Civil Engineering, Katholieke Universiteit Leuven,
Celestijnenlaan 131, B-3001 Heverlee, Belgium

(Received 19 November 1996; accepted for publication 11 February 1997)

Vibration attenuation at junctions of point connected plates is an important aspect of structure-borne sound transmission in many realistic structures. The vibrational response of point connected plates can be predicted using statistical energy analysis (SEA), where coupling loss factors are calculated based on mobility formulations. However, in the process of deriving the coupling loss factors for point connected systems, a number of assumptions and simplifications were introduced. In this paper, a more advanced calculation model, based on a wave approach for elastically coupled semi-infinite plates, is presented. Local rigid connections are modeled using an elastic interlayer characterized by a space-dependent stiffness. Periodic boundary conditions are assumed at the junction and the plate response is described using a Fourier decomposition technique. Calculation results obtained for two different junction geometries demonstrate the influence of the distance between the point connections on the structure-borne sound transmission. Finally, a comparison between a number of numerical and experimental results illustrates the ability of the model to predict the influence of the density and the width of the point connections with good accuracy.

© 1997 Acoustical Society of America. [S0001-4966(97)02906-8]

PACS numbers: 43.40.Dx, 43.40.At [CBB]

LIST OF SYMBOLS

| | | | |
|-----------------|--|--------------|--|
| A, B | bending wave amplitudes | ζ | displacement in z direction |
| C_g | group velocity | η | displacement in y direction |
| e_p | eccentricity of plate edge with respect to the junction line | η_{ii} | internal loss factor of subsystem i |
| f^n | cut-on frequency of wave component n | η_{ij} | coupling loss factor describing the energy flow between subsystems i and j |
| F_x, F_y, F_z | force per unit width in x, y , and z directions | ξ | displacement in x direction |
| I_{px} | energy flow per unit width in x direction | θ_i | angle with x axis of incident wave |
| I_{px} | space averaged intensity in x direction | θ_p | angle characterizing plate orientation |
| k | wave number | τ | angle-dependent transmission coefficient |
| K_x, K_y, K_z | spring stiffness in x, y , and z directions | $\bar{\tau}$ | angle-averaged transmission coefficient |
| K_{xmn} | complex stiffness coefficient in x direction | Φ | potential function describing quasi-longitudinal wave propagation |
| K_{rz} | rotational spring stiffness around z axis | Ψ | scalar stream function describing in-plane transverse wave propagation |
| l_{ij} | effective junction length between plates i and j | ω | circular frequency |
| L | spatial period of the boundary conditions | $()^*$ | complex conjugate of (complex number) |
| L_{vij} | velocity level difference between plate i and j | Subscripts | |
| M_z | moment per unit width around the z axis | 0 | referring to junction beam |
| n | order of wave component | 1,2 | referring to plates 1 and 2 |
| N | maximum order of wave component | i | referring to incident wave |
| N_c | number of point connections | l | referring to elastic layer |
| N_p | number of plates | m | referring to equation number |
| $\text{Re}()$ | real part of (complex number) | n | referring to wave component n |
| S | surface area | p | referring to plate number |
| t | time | x, y, z | referring to x, y , and z direction |
| x, y, z | coordinates | B, L, T | referring to bending, quasi-longitudinal, and in-plane transverse waves |
| α_z | rotation around z axis | | |

INTRODUCTION

The application of elastic interlayers as vibration isolators at junctions between walls and slabs is reported by several authors as a key solution in reducing flanking transmis-

sion in buildings.^{1,2} Although the overall sound insulation would improve significantly by applying elastic layers to all junctions in a building, the stability requirements of the construction imply that only a limited number of walls qualify to

be elastically coupled. Since vibrational energy is redirected rather than dissipated in such junctions,³ locating elastic layers at only some of the junctions can have a negative effect on sound insulation. One solution to these problems is to use local rigid connections at distinct positions along the plate junction. The aim of this solution is to extend the applicability of elastic junctions in buildings by improving the mechanical stability of the structure, without exceedingly compromising the benefits of the elastic layer. In this study, a calculation model based on the work of Mees and Vermeir¹ is developed to estimate the effect of local rigid connections on the vibrational energy transmission. This model represents an extension of existing models for structure-borne sound transmission between semi-infinite plates and it provides a more advanced procedure of calculating coupling loss factors for statistical energy analysis (SEA).

The problem of predicting structure-borne sound transmission between semi-infinite plates has been solved by several authors. Cremer *et al.*⁴ investigated longitudinal and bending wave transmission at normal incidence for different junctions, as well as bending wave transmission for oblique incidence at a corner junction. The bending wave transmission at symmetric cross junctions for random incidence was treated by Kihlman.⁵ More advanced calculation models, predicting the transmission coefficients for bending, quasi-longitudinal, and in-plane transverse waves for random incidence, at an arbitrary L , T , or cross junction, were developed by Wöhle, Beckmann, and Schreckenbach^{6,7} and by Craven and Gibbs.^{8,9} Cremer *et al.*⁴ and Wöhle *et al.*⁶ studied the possibility of introducing elastic interlayers at junctions of plates, but a more detailed investigation of this problem was carried out by Mees and Vermeir.¹ In the latter work, the effect of thickness resonances was taken into account by modeling the elastic interlayer as a wave supporting medium. Further, the authors estimated the apparent stiffness of the interlayer using finite element simulations.

The vibrational energy flow between point connected structures has been studied by Lyon *et al.*¹⁰ and Manning¹¹ in the context of SEA. In the work performed by both authors, an expression for the coupling loss factor for single point connected systems is derived based on a mobility function formulation. In the latter analysis, multi-component coupling, i.e., coupling between the different degrees of freedom involved at the junction, is neglected. For multi-point connected systems, an additional simplification is introduced when the different points are assumed to act independently. As a result, the effect of each point can be treated individually and the coupling loss factor derived for a single point connection simply needs to be multiplied by the number of points to estimate the energy flow between multi-point coupled systems.¹² At frequencies where the distance between the point connections is small in comparison to the wavelength, however, the motion of the different point connections is highly correlated, leading to large discrepancies between measurement and prediction. A practical solution of this problem is to treat a junction of point connected plates as a line junction at low frequencies.

In the model presented in this paper, multi-component coupling is included in the analysis, and no assumptions

have been introduced concerning the interaction between the different point connections. In a number of numerical simulations, part of the simplifications commonly used in SEA are verified by studying the results for different cases of point spacing. Finally, the calculation model is validated by comparing measured and predicted velocity level differences between point connected plates.

In this study, local rigid connections are modeled by assuming a finite contact length. Since in the frequency range under consideration the wavelength is considerably larger than the contact length, local rigid connections will be referred to as “point connections” in the following text.

I. CALCULATION MODEL

The theoretical model for predicting structure-borne sound transmission between point connected plates is based on a calculation model for elastically coupled plates.¹ Since various aspects of the work performed in Ref. 1 have been incorporated in the present research, only the most essential topics of the existing theory will be reviewed in the next paragraphs.

A. Rigid junction

The study presented in this paper is restricted to linear elastic, isotropic, homogeneous plates with a small thickness compared to the bending wavelength. Consequently, the effect of rotatory inertia and shear deformation on the bending wave propagation is neglected. Although floors and walls cannot be regarded as thin over the whole frequency range of interest in building acoustics, it is assumed that thin plate theory can be used to study the effect of local rigid connections on structure-borne sound transmission. Internal damping of the plate material was not included in the analysis. Three wave types were taken into account to characterize the energy flow at the plate junction: bending, quasi-longitudinal, and in-plane transverse waves.

The excitation is modeled as a plane wave traveling toward the junction with oblique incidence. This wave generates bending and in-plane waves propagating away from the junction on all plates. The amplitudes of the traveling waves follow from equilibrium and continuity conditions at the junction line. The energy flow through the junction is quantified by reflection and transmission coefficients, which are defined as the ratios of the reflected and transmitted intensities, respectively, to the intensity carried by the incident wave. The transmission coefficient for random incidence is calculated by integrating over all angles of incidence.

The geometry considered in this study is illustrated in Fig. 1. The junction consists of an assembly of semi-infinite plates coupled by a massless and perfectly flexible beam, which is not restricted with respect to its translation. The junction beam is characterized by an offset e_p , by which a plate is mounted at the junction. Two to four plates can be connected to the junction beam at any arbitrary coupling angle. The plate orientation is determined by the angle θ_p between the plate and the x - z plane of the global coordinate system. The displacements ξ_p , η_p , and ζ_p , and the rotation α_{zp} on the plate edge are expressed in the local coordinate

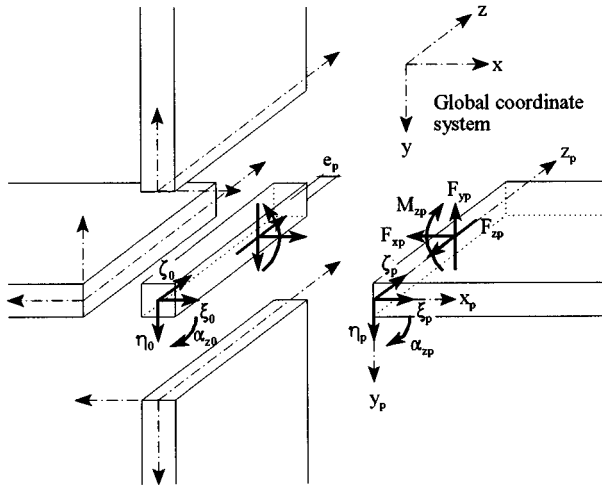


FIG. 1. Geometry of a rigid plate junction.

system. The forces F_{xp} , F_{yp} , and F_{zp} , and moment M_{zp} can be derived from the displacements using elementary formulae of mechanics.

For the case of a rigid junction, the plate edge is rigidly connected to the junction beam. Consequently, the displacements at the plate edge are equal to the displacements of the junction beam, leading to four continuity conditions for each plate (expressed in the local coordinate system):¹

$$\xi_p = \xi_0 \cos \theta_p + \eta_0 \sin \theta_p, \quad (1)$$

$$\eta_p = -\xi_0 \sin \theta_p + \eta_0 \cos \theta_p + e_p \alpha_{z0}, \quad (2)$$

$$\zeta_p = \zeta_0, \quad (3)$$

$$\alpha_{zp} = \alpha_{z0}. \quad (4)$$

Since the junction beam is modeled as massless and perfectly flexible, wave propagation is not supported by this element. Consequently, the sum of forces acting on the junction beam equals zero and the following four equilibrium conditions can be expressed in the global coordinate system:¹

$$\sum_p (F_{xp} \cos \theta_p - F_{yp} \sin \theta_p) = 0, \quad (5)$$

$$\sum_p (F_{xp} \sin \theta_p + F_{yp} \cos \theta_p) = 0, \quad (6)$$

$$\sum_p F_{zp} = 0, \quad (7)$$

$$-\sum_p M_{zp} + \sum_p e_p F_{yp} = 0. \quad (8)$$

Equation (8) shows that, by the presence of the junction offset e_p , the transverse force F_{yp} contributes to the total bending moment acting on the junction beam. In addition, the transverse displacement η_p on the plate edge is related to the rotation of the junction beam α_{z0} , as stated in Eq. (2). For a rigid plate junction, the introduction of a junction offset is not entirely justified, since the transition from one plate to another is not clearly defined. Consequently, the offset is taken to be zero for a rigid junction. For a junction of plates

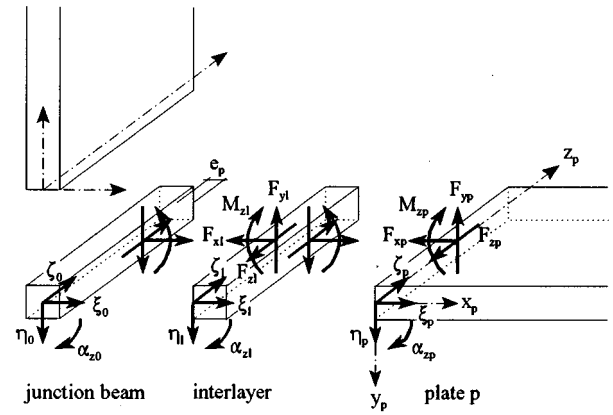


FIG. 2. Geometry at a junction of elastically coupled plates.

coupled by a beam¹³ or an elastic interlayer however, the plate edge can be easily located, since a clear eccentricity of the plate edge with respect to the junction line has been introduced. Further, a comparison between measurement and calculation, as performed by Craik and Osipov,² illustrated the importance of including the junction offset in the analysis for elastically coupled plates.

B. Elastically connected plates

As a first step in modeling point connections, an isotropic, homogeneous elastic interlayer is introduced at the plate junction, as illustrated in Fig. 2. The effect of the interlayer is taken into account by a transformation of forces and displacements on the plate edge to the corresponding forces and displacements on the edge of the interlayer. This transformation depends on the dynamic properties of the interlayer and on its mathematical description.

At frequencies where the interlayer thickness is small compared to the longitudinal and transverse wavelengths of the interlayer material, the elastic layer can be approximated by a spring dash-pot system. In this case, the forces acting on both sides of the interlayer are equal:

$$F_{xl} = F_{xp}, \quad (9a)$$

$$F_{yl} = F_{yp}, \quad (9b)$$

$$F_{zl} = F_{zp}, \quad (9c)$$

$$M_{zl} = M_{zp}. \quad (9d)$$

The displacements at the edge of the interlayer are related to the corresponding plate edge displacements by four spring constants, which characterize the elastic junction:

$$\xi_l = \xi_p - \frac{F_{xp}}{K_x}, \quad (10a)$$

$$\eta_l = \eta_p - \frac{F_{yp}}{K_y}, \quad (10b)$$

$$\zeta_l = \zeta_p - \frac{F_{zp}}{K_z}, \quad (10c)$$

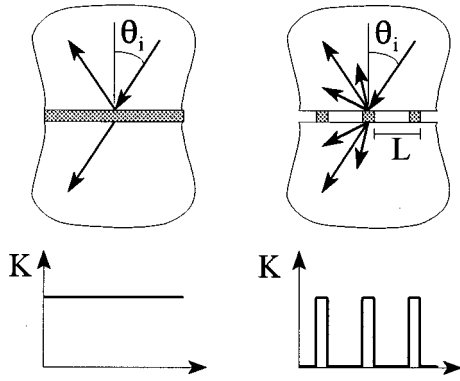


FIG. 3. Point connections are modeled by a periodically varying interlayer stiffness, resulting in diffusely transmitted structure-borne sound.

$$\alpha_{zl} = \alpha_{zp} + \frac{M_{zp}}{K_{rz}}. \quad (10d)$$

The positive sign in Eq. (10d) follows from the conventions shown in Fig. 2. For some idealized cases of interlayer deformation, the spring constants K_x , K_y , K_z , and K_{rz} can be derived from the material properties and dimensions of the interlayer.¹ The internal damping of the interlayer material is taken into account by a complex Young's modulus.

Thickness resonances can be observed in the interlayer at frequencies where the interlayer thickness is large compared to shear wavelength in the interlayer material. A more exact theory,¹ which includes wave propagation in the interlayer, incorporates the influence of these resonance effects. In this paper however, the simple spring dash-pot system is adopted.

C. Periodically connected plates

1. Scattered wave fields

A junction of plates coupled by local rigid connections is modeled by means of an elastic layer with a space-dependent stiffness. Point connections are simulated by a local, stepwise increase of the interlayer stiffness, as illustrated in Fig. 3. The interlayer stiffness is assumed to vary periodically, leading to periodic boundary conditions for the connected semi-infinite plates. By analogy with diffuse reflection of air-borne sound from a periodic structure of a wall,¹⁴ the plate response is described by a scattered wave field.

The excitation of the junction of semi-infinite plates is modeled as a plane, unit amplitude wave traveling toward the junction line. For the case of an incident plane wave characterized by a wave number k_i and an angle of incidence θ_i , the resulting bending wave field on each plate is described by the following expression (the time dependence $e^{j\omega t}$ will be omitted in the following equations for reasons of readability):

$$\eta_p(x, z) = \sum_n (A_{pn} e^{-jk_{Bxn1}x} + B_{pn} e^{-jk_{Bxn2}x}) e^{-j(k_i \sin \theta_i + 2n\pi/L)z}, \quad (11)$$

where the summation is carried out for n ranging from $-\infty$ to $+\infty$. Equation (11) states that the transverse displacement is the result of the superposition of a number of plane waves traveling away from the junction. These waves are characterized by a z dependence, which is imposed by the incident wave and the spatial period L of the boundary conditions. Substitution of Eq. (11) in the bending wave equation for thin isotropic plates, yields the wave numbers k_{Bxn1} and k_{Bxn2} :

$$k_{Bxn1} = \sqrt{k_B^2 - (k_i \sin \theta_i + 2\pi n/L)^2}, \quad (12)$$

$$k_{Bxn2} = -j\sqrt{k_B^2 + (K_i \sin \theta_i + 2\pi n/L)^2}. \quad (13)$$

The wave number k_{Bxn1} can be either real or imaginary. A real wave number corresponds to a traveling wave, whereas an imaginary wave number indicates an exponentially decaying nearfield. Since k_{Bxn2} is always imaginary, the contribution of bending wave component n to the resulting energy flow at the junction is determined by the first part of Eq. (11).

The quasi-longitudinal and in-plane transverse waves can be expressed in terms of a potential Φ_p and a scalar stream function Ψ_p .⁴ These functions are related to the in-plane displacements ξ_p and ζ_p by

$$\xi_p = \frac{\partial \Phi_p}{\partial x} - \frac{\partial \Psi_p}{\partial z}, \quad \zeta_p = \frac{\partial \Phi_p}{\partial z} + \frac{\partial \Psi_p}{\partial x}. \quad (14)$$

Based on these definitions, the equations of motion for in-plane waves can be rewritten into two uncoupled differential equations, corresponding to, respectively, quasi-longitudinal and in-plane transverse wave propagation:

$$(\Delta + k_L^2)\Phi_p = 0, \quad (\Delta + k_T^2)\Psi_p = 0. \quad (15)$$

The diffuse reflection and transmission of in-plane waves at a periodic junction of semi-infinite plates is described by

$$\Phi_p(x, z) = \sum_n \Phi_{pn} e^{-jk_{Lxn}x} e^{-j(k_i \sin \theta_i + 2\pi n/L)z}, \quad (16)$$

$$\Psi_p(x, z) = \sum_n \Psi_{pn} e^{-jk_{Txn}x} e^{-j(k_i \sin \theta_i + 2\pi n/L)z}. \quad (17)$$

The in-plane wave numbers k_{Lxn} and k_{Txn} follow from substitution into Eqs. (15), and expressions similar to Eq. (12) can be derived.

The cut-on frequency of wave component n : f_{IT}^n (the indices I and T refer to the incident and transmitted wave type, respectively) is defined as the lowest frequency at which the wave component transforms from a near field into a propagating wave. It can be shown using Eq. (12) that f_{IT}^n increases with increasing n . Consequently, only the primary wave components ($n=0$) represent energy carrying waves at low frequencies. Above their respective cut-on frequencies, higher-order wave components ($n \neq 0$) contribute to the energy flow at the plate junction, resulting in diffusely reflected and transmitted structure-borne sound. Although the theoretically exact bending and in-plane wave fields are composed of an infinite number of wave components, the

summations in Eqs. (11), (16), and (17) are carried out for $n = -N \cdots +N$, where N is determined based on convergence criteria.

On the plate edge $x=0$, the displacements and forces can be expressed in the form

$$X_p(z) = \sum_n X_{pn} e^{-j[k_i \sin \theta_i + 2n\pi/L]z},$$

$$X = \xi, \eta, \zeta, \alpha_z, F_x, F_y, F_z, \text{ and } M_z. \quad (18)$$

The complex amplitudes X_{pn} are derived from Eqs. (11), (16), and (17), based on the formulae of mechanics. The complex amplitudes corresponding to bending wave propagation (η_{pn} , α_{zpn} , F_{ypn} , and M_{zpn}) are calculated as a linear combination of A_{pn} and B_{pn} , whereas the complex amplitudes corresponding to in-plane wave motion (ξ_{pn} , ζ_{pn} , F_{xpn} , and F_{zpn}) are expressed in terms of Φ_{pn} and Ψ_{pn} . The displacements of the junction beam are characterized by the same z dependence as the displacements on the plate edge. Consequently, expressions similar to Eq. (18) are applicable for the displacements of the junction beam:

$$X_0(z) = \sum_n X_{0n} e^{-j(k_i \sin \theta_i + 2n\pi/L)z},$$

$$X = \xi, \eta, \zeta, \text{ and } \alpha_z. \quad (19)$$

The complex amplitudes ξ_{0n} , η_{0n} , ζ_{0n} , and α_{z0n} , together with A_{pn} , B_{pn} , Φ_{pn} , and Ψ_{pn} represent the unknowns of the problem. The continuity and equilibrium conditions at the junction line lead to a set of linear equations, the solution of which yields the unknown amplitudes.

2. Continuity conditions

The introduction of an interlayer with a periodically varying spring stiffness requires a more complex formulation of the boundary conditions as for the case of a homogeneous interlayer. In this section, the continuity condition for the displacement in the x direction of the local coordinate system is derived. Similar expressions can be obtained for the remaining degrees of freedom.

For elastically connected plates, Eqs. (1)–(4) need to be modified to account for the fact that continuity is required between the displacements of the junction beam and the corresponding displacements at the edge of the interlayer (and not at the edge of the plate). Consequently, Eq. (1) can be rewritten as

$$\xi_l(z) = \xi_0(z) \cos \theta_p + \eta_0(z) \sin \theta_p. \quad (20)$$

By combining Eqs. (10a) and (20), and expressing forces and displacements as stated by Eqs. (18) and (19), the continuity condition for displacement in the x direction is found as

$$K_x(z) \sum_{n=-N}^{+N} (\xi_{pn} - \xi_{0n} \cos \theta_p - \eta_{0n} \sin \theta_p) e^{-j(2\pi n/L)z} - \sum_{n=-N}^{+N} F_{xpn} e^{-j(2\pi n/L)z} = 0. \quad (21)$$

In this equation, the spring stiffness K_x represents an arbitrary, periodical function of z . Equation (21) is now multi-

plied by $e^{+j(2\pi m/L)z}$ and integrated over one spatial period of the boundary, leading to the following expression:

$$\sum_{n=-N}^{+N} K_{xmn} (\xi_{pn} - \xi_{0n} \cos \theta_p - \eta_{0n} \sin \theta_p) - L F_{xpm} = 0, \quad (22)$$

where

$$K_{xmn} = \int_0^L K_x(z) e^{-j[2\pi(n-m)/L]z} dz. \quad (23)$$

By repeating this procedure for values of m ranging from $-N$ to $+N$, the continuity for the x direction is ensured by $2N+1$ equations per plate. The summation in Eq. (22) implies that the amplitudes of all wave components are coupled by the continuity conditions. Only when the spring stiffness is independent of z , as for a homogeneous elastic interlayer, the amplitudes will be uncoupled, since, for this case, the complex stiffness coefficients K_{xmn} equal zero for $m \neq n$.

3. Equilibrium of the junction beam

The equilibrium conditions of the junction beam are derived using an approach similar to the formulation of the continuity conditions. The equilibrium in the x direction of the global coordinate system is found by rearranging Eqs. (5), (9a), and (9b), and expressing the forces at the plate edge as illustrated in Eq. (18):

$$\sum_{p=1}^{N_p} \sum_{n=-N}^{+N} (F_{xpn} \cos \theta_p - F_{ypn} \sin \theta_p) e^{-j(2\pi n/L)z} = 0, \quad (24)$$

where N_p represents the number of plates coupled at the junction. After multiplication of Eq. (24) by $e^{+j(2\pi m/L)z}$ and the subsequent integration over the spatial period L of the boundary conditions, the equilibrium condition can be formulated for each wave component separately:

$$\sum_{p=1}^{N_p} (F_{xpm} \cos \theta_p - F_{ypm} \sin \theta_p) = 0, \quad m = -N \cdots +N, \quad (25)$$

leading to $2N+1$ equations. Similar expressions are obtained for the equilibrium conditions corresponding to the remaining degrees of freedom of the junction beam.

4. Evaluation of the transmission coefficient

The intensity normal to the junction line carried by the scattered bending waves can be expressed in terms of the forces and displacements at the plate edge $x=0$ (Fig. 1) as follows:⁴

$$I_{Bpx}(z) = \frac{1}{2} \operatorname{Re}(M_{zp}(z)(j\omega\alpha_{zp}(z))^* + F_{yp}(z)(-j\omega\eta_p(z))^*). \quad (26)$$

The bending wave intensity I_{Bpx} in Eq. (26) is a periodic function of z , since the energy flow is dependent on the stiffness of the interlayer. For a junction of point connected plates, the intensity equals zero at the uncoupled regions, and the energy flow takes place exclusively at the point connec-

tions. Since a space-dependent transmission coefficient is impractical, Eq. (26) is averaged over one spatial period L by performing the following integration:

$$\overline{I_{Bpx}} = \frac{1}{L} \int_0^L I_{Bpx}(z) dz. \quad (27)$$

Equation (27) can be rewritten using Eqs. (18) and (26), leading to

$$\overline{I_{Bpx}} = \frac{1}{2} \operatorname{Re} \left(\sum_n (M_{zpn}(j\omega\alpha_{zpn})^* + F_{ypn}(-j\omega\eta_{pn})^*) \right). \quad (28)$$

From Eq. (28) one can conclude that averaging the bending wave intensity over one spatial period is equivalent to adding the intensities carried by the individual wave components of the scattered wave field. As a consequence, the summation in Eq. (28) should include wave components representing traveling waves only.

The displacements and forces corresponding to in-plane wave motion are composed of contributions from quasi-longitudinal and transverse waves. Consequently, the displacement in the x direction at the plate edge $x=0$ for wave component n , can be expressed as

$$\xi_{pn} = \xi_{pn\Phi} + \xi_{pn\Psi}, \quad (29)$$

where the indices Φ and Ψ refer to quasi-longitudinal and transverse waves, respectively. Similar decompositions are applicable for ζ_{pn} , F_{xpn} , and F_{zpn} . As a result, the space averaged intensities in the x direction, caused by transmitted quasi-longitudinal and in-plane transverse waves, are given by

$$\overline{I_{Lpx}} = \frac{1}{2} \operatorname{Re} \left(\sum_n (F_{xpn\Phi}(-j\omega\xi_{pn\Phi})^* + F_{zpn\Phi}(-j\omega\zeta_{pn\Phi})^*) \right), \quad (30)$$

$$\overline{I_{Tpx}} = \frac{1}{2} \operatorname{Re} \left(\sum_n (F_{xpn\Psi}(-j\omega\xi_{pn\Psi})^* + F_{zpn\Psi}(-j\omega\zeta_{pn\Psi})^*) \right). \quad (31)$$

The space-averaged intensities in Eqs. (28), (30), and (31) depend on the propagation direction of the incident wave. The angle-dependent transmission coefficient is calculated as

$$\tau_p(\theta_i) = \frac{\overline{I_{px}}(\theta_i)}{I_{ix}(\theta_i)}, \quad (32)$$

where I_{ix} represents the intensity in the x direction carried by the incident wave. Finally, the transmission coefficient for random incidence is calculated by averaging $\tau_p(\theta_i)$ over all angles of incidence:⁴

$$\overline{\tau_p} = \int_0^{\pi/2} \tau_p(\theta_i) \cos \theta_i d\theta_i. \quad (33)$$

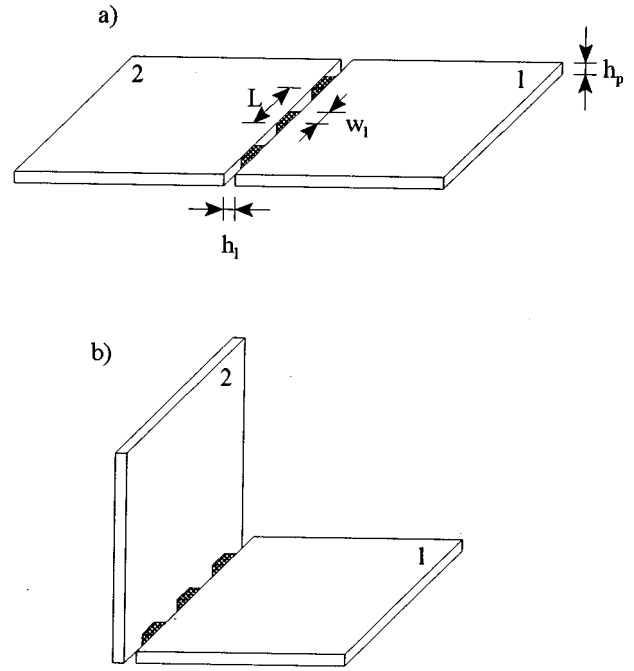


FIG. 4. Numerical simulations were carried out for two different junction geometries: an in-line junction (a) and a corner (b) junction of two concrete plates.

II. NUMERICAL RESULTS

The theory outlined in Sec. I C is generally applicable for any periodic variation of the interlayer stiffness. The numerical results presented in this chapter, however, are restricted to periodic plate junctions where only one rigid connection is positioned within a spatial period. Consequently, the local rigid connections are assumed to be equally spaced.

Calculations were performed for two different junction geometries as illustrated in Fig. 4: an in-line (a) and a corner junction (b). The material properties of the plates and the elastic interlayer are: Young's modulus $E = 4.0 \times 10^{10}$ Pa, Poisson ratio $\nu = 0.17$, density $\rho = 2500$ kg/m³. The internal loss of the interlayer material is not included. The plate thickness h_p is 0.15 m and the local rigid connections have the following dimensions (Fig. 4): width $w_1 = 0.15$ m and height $h_1 = 0.015$ m. The junction offset was taken as half the plate thickness in all cases. At the point connections, the spring stiffness for the different degrees of freedom are calculated as¹

$$K_x = \frac{E(1-\nu)}{(1-2\nu)(1+\nu)} \frac{h_p}{h_1}, \quad (34a)$$

$$K_y = \frac{E}{2(1+\nu)} \frac{h_p}{h_1}, \quad (34b)$$

$$K_z = \frac{E}{2(1+\nu)} \frac{h_p}{h_1}, \quad (34c)$$

$$K_{rz} = \frac{E(1-\nu)}{(1-2\nu)(1+\nu)} \frac{h_p^3}{12h_1}. \quad (34d)$$

TABLE I. Five different cases of point spacing depending on the ratio of the spatial period to the bending wavelength at 100 Hz, and the corresponding cut-on frequencies f_{BB}^{-1} , f_{LL}^{-1} , and f_{TT}^{-1} at which the spatial period equals half the wavelength of the corresponding wave type.

| Case | $L/\lambda_{B,100}$ Hz | f_{BB}^{-1} [Hz] | f_{LL}^{-1} [Hz] | f_{TT}^{-1} [Hz] |
|------|------------------------|--------------------|--------------------|--------------------|
| 1 | 1/16 | 6400.0 | 9771.6 | 6294.9 |
| 2 | 1/8 | 1600.0 | 4885.8 | 3147.4 |
| 3 | 1/4 | 400.0 | 2442.9 | 1573.7 |
| 4 | 1/2 | 100.0 | 1221.4 | 786.9 |
| 5 | 1 | 25.0 | 610.7 | 393.4 |

These spring constants correspond to an idealized case of interlayer deformation, since it is assumed that the interlayer displacements do not vary with y or z . Other idealized cases may be considered,¹ but for the junctions studied in this chapter, preliminary simulations showed that the energy flow through the junction is rather insensitive to the assumed interlayer deformation.

The aim of the numerical simulations is to illustrate the effect of the distance between the point connections on the structure-borne sound transmission. For this purpose, five different cases of point spacing were considered for each junction geometry. The spatial period L was chosen as a function of the bending wavelength of the plates at 100 Hz, as listed in Table I. The calculations were carried out in small frequency steps (1/27th octave) and then averaged over each one-third octave band. The results are presented in terms of the transmission loss R_p , which is defined as

$$R_p = -10 \log(\overline{\tau_p}). \tag{35}$$

A. In-line junction

The structure-borne sound transmission for this type of plate junction is characterized by the absence of wave conversion between bending and in-plane waves. The bending wave transmission loss for the five cases of point spacing is shown in Fig. 5. A nearly perfect transmission, corresponding to the case of a rigid connection, can be observed up to a certain frequency. This frequency coincides with the cut-on frequency of wave component $n = -1$ for incident and transmitted bending waves: f_{BB}^{-1} . The rapid increase of the attenuation above this frequency illustrates the fact that an in-line junction of point connected plates can be regarded as a low-pass filter with respect to structure-borne transmission.

For the case of transmission between identical plates and for incident and transmitted waves of the same wave type, it can be shown using Eq. (12) that the cut-on frequency of wave component $n = -1$ can be calculated as the frequency at which the spatial period L equals half the wavelength of the considered wave type. These frequencies are listed in Table I for the five cases of point spacing and for the three wave types included in the analysis. The influence of wave component $n = -1$ on the bending wave transmission is illustrated in Fig. 6, showing the transmission coefficient as a function of the angle of incidence for three different frequencies. The results are presented for case 3, which is characterized by a cut-on frequency $f_{BB}^{-1} = 400$ Hz. While the energy flow at 315 Hz approximates the case of perfect transmis-

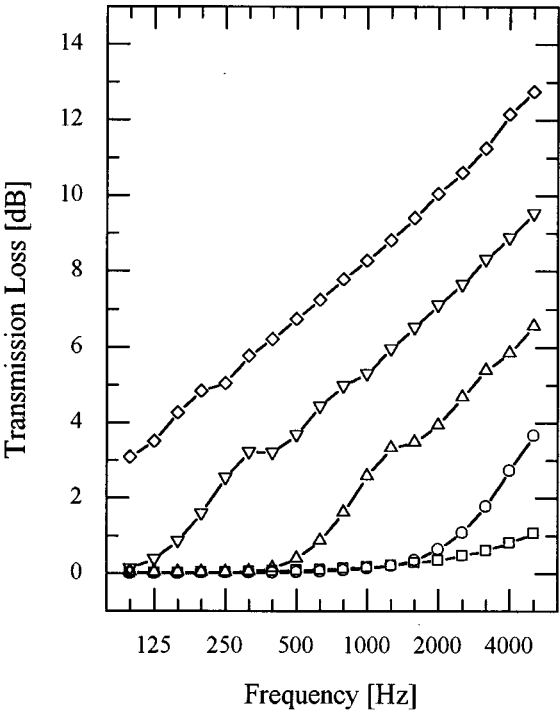


FIG. 5. Transmission loss for incident and transmitted bending waves at the in-line junction for five cases of point spacing: 1 (\square), 2 (\circ), 3 (\triangle), 4 (∇), and 5 (\diamond).

sion, the results above f_{BB}^{-1} exhibit a local minimum at the cut-on angle of wave component $n = -1$. The sudden increase of transmission loss in Fig. 5 can be explained by considering that, with increasing frequency, the local minimum of the angle-dependent transmission coefficient in Fig.

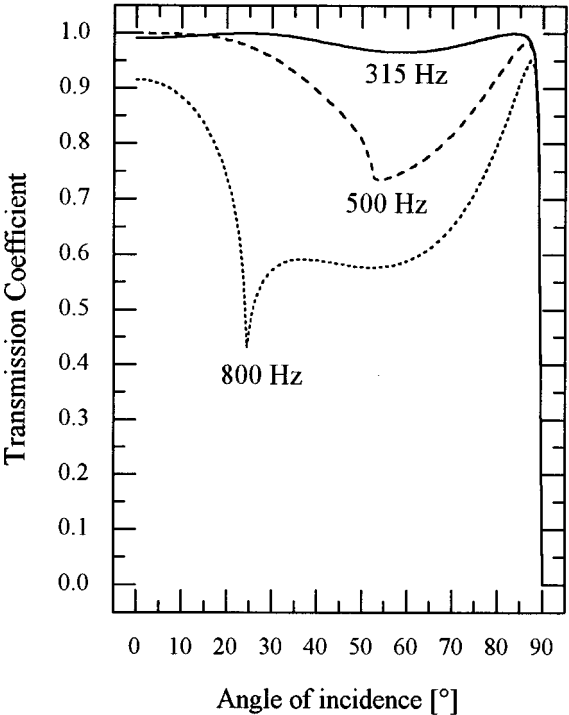


FIG. 6. Transmission coefficient as a function of the angle of incidence for incident and transmitted bending waves at the in-line junction, for case 3 of point spacing and three frequencies: 315 Hz (—), 500 Hz (---), and 800 Hz (·····).

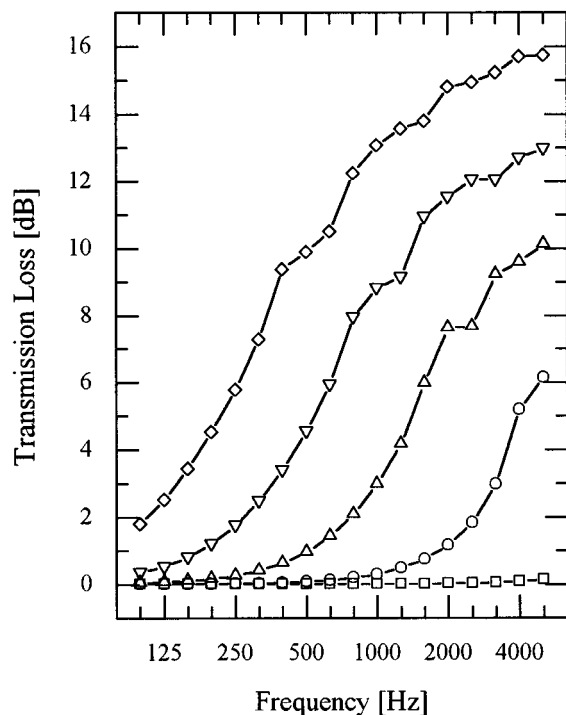


FIG. 7. Transmission loss for incident and transmitted quasi-longitudinal waves at the in-line junction for five cases of point spacing: 1 (□), 2 (○), 3 (△), 4 (▽), and 5 (◇).

6 becomes more pronounced and its location is shifted toward smaller angles of incidence. Consequently, as the frequency increases, gradually more energy will be carried by the reflected bending wave component $n = -1$.

The transmission loss for incident and transmitted quasi-longitudinal waves is shown in Fig. 7. Again a low-pass filter effect can be observed, but the frequency at which the transmission loss starts to increase is not clearly defined. The transmission loss curve displays a small “dip” in the one-third octave band containing the cut-on frequency f_{LL}^{-1} (Table I), but the transmission loss at this frequency remains well above 0 dB. While the increasing transmission loss above f_{LL}^{-1} is governed by mechanisms similar to the ones discussed for the case of bending wave transmission, the increasing transmission loss below f_{LL}^{-1} is difficult to interpret, but may be related to wave conversion between the in-plane waves. The relatively important coupling between quasi-longitudinal and in-plane transverse waves is demonstrated by the fact that the reflected energy is almost equally distributed between both types of in-plane waves, as illustrated in Fig. 8.

The results for incident and transmitted in-plane transverse waves are presented in Fig. 9. Apart from the different positions of the local minima in the transmission loss curve, a strong similarity between Fig. 9 and Fig. 7 can be observed. Compared to the results for longitudinal wave transmission, the local minima are shifted towards lower frequencies due to the lower cut-on frequencies f_{TT}^{-1} .

B. Corner junction

At a right-angled plate junction, the bending and in-plane wave motion are coupled, and the structure-borne

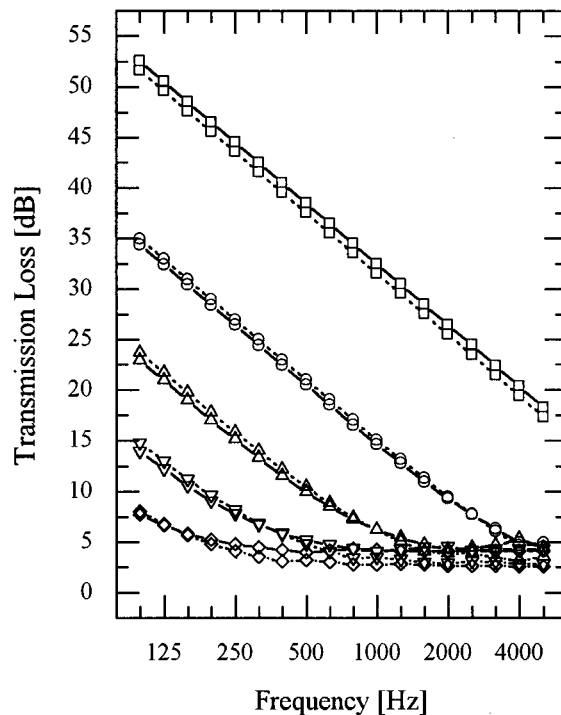


FIG. 8. Transmission loss for incident quasi-longitudinal wave, reflected quasi-longitudinal (—) and in-plane transverse (·····) waves. In-line junction, five cases of point spacing: 1 (□), 2 (○), 3 (△), 4 (▽), and 5 (◇).

sound transmission is governed by the complex mechanism of wave conversion in the entire frequency range. In Figs. 10–14 the transmission loss for the different cases of point spacing are presented and the result for a continuous elastic

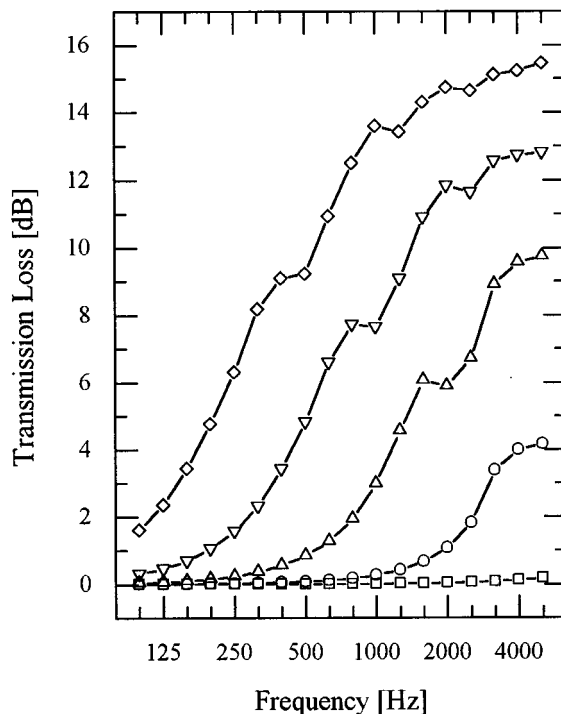


FIG. 9. Transmission loss for incident and transmitted in-plane transverse waves at the in-line junction for five cases of point spacing: 1 (□), 2 (○), 3 (△), 4 (▽), and 5 (◇).

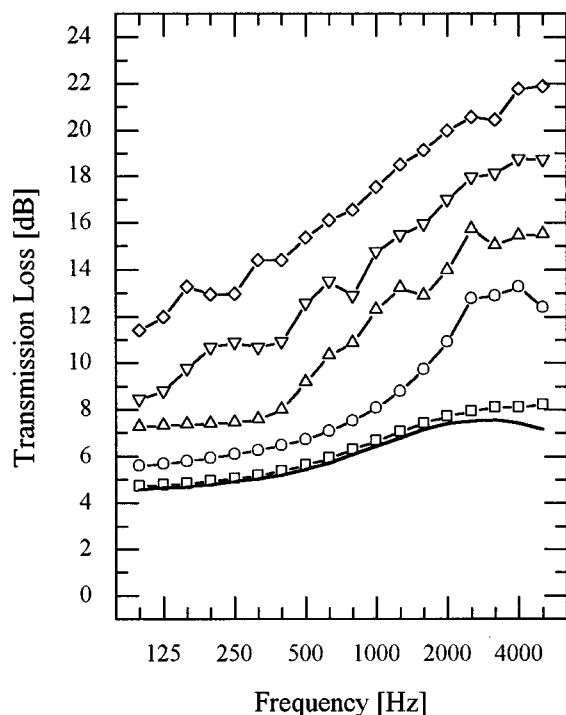


FIG. 10. Transmission loss for incident and transmitted bending waves at the corner junction for five cases of point spacing: 1 (□), 2 (○), 3 (△), 4 (▽), and 5 (◇). Reference curve: continuous elastic interlayer (—).

interlayer is taken as a reference. The structure-borne sound excitation is applied to plate 1, which is periodically coupled to plate 2 as shown in Fig. 4(b).

The transmission loss corresponding to incident and

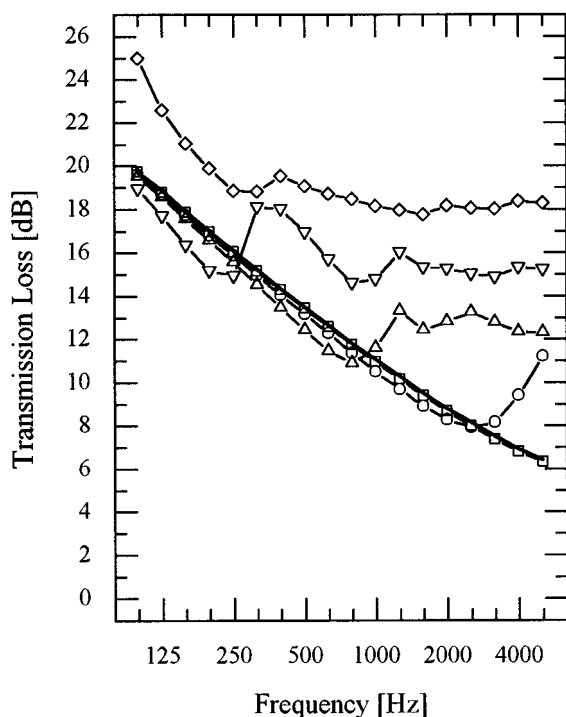


FIG. 11. Transmission loss for incident bending wave and transmitted quasi-longitudinal wave at the corner junction for five cases of point spacing: 1 (□), 2 (○), 3 (△), 4 (▽), and 5 (◇). Reference curve: continuous elastic interlayer (—).

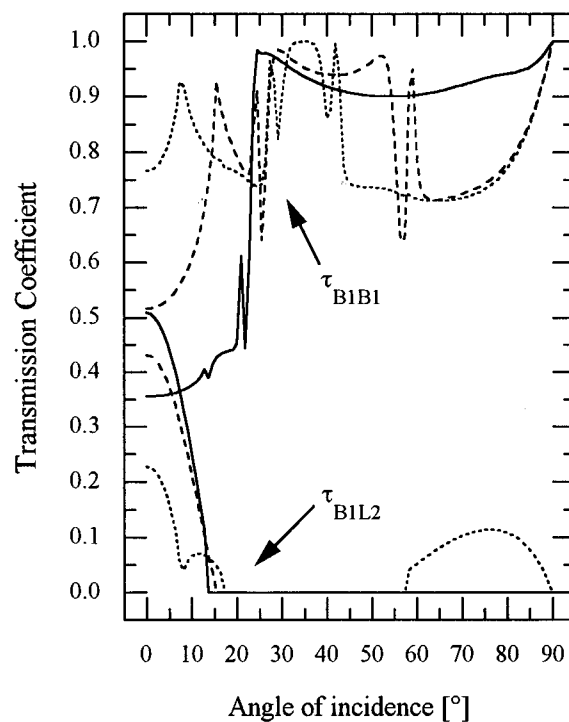


FIG. 12. Transmission coefficient as a function of the angle of incidence. Incident bending wave, reflected bending (τ_{B1B1} , upper curves) and transmitted quasi-longitudinal waves (τ_{B1L2} , lower curves). Corner junction, case 3 of point spacing, three different frequencies: 800 Hz (—), 1000 Hz (---), and 1250 Hz (·····).

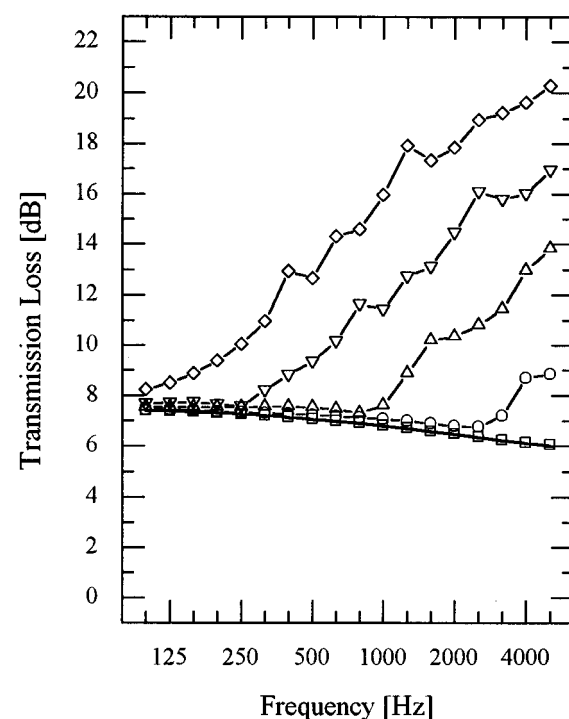


FIG. 13. Transmission loss for incident and transmitted quasi-longitudinal waves at the corner junction for five cases of point spacing: 1 (□), 2 (○), 3 (△), 4 (▽), and 5 (◇). Reference curve: continuous elastic interlayer (—).

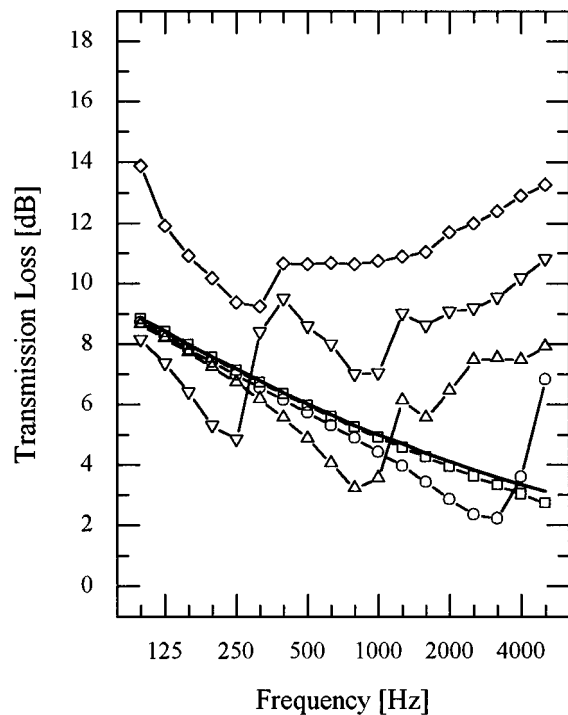


FIG. 14. Transmission loss for incident quasi-longitudinal wave and transmitted bending wave at the corner junction for five cases of point spacing: 1 (\square), 2 (\circ), 3 (\triangle), 4 (∇), and 5 (\diamond). Reference curve: continuous elastic interlayer (—).

transmitted bending waves is plotted in Fig. 10. As for an in-line junction, the bending wave attenuation increases with increasing spatial period. The cut-on frequencies f_{BB}^{-1} can be identified in one-third octave bands above which the slope of the transmission loss curves tends to increase. However, only for case 1, corresponding to the highest density of point connections, the result for the continuous elastic junction is approximated. The transmission loss for the remaining cases of point spacing exceeds the reference curve in the entire frequency range. This observation leads to the conclusion that even if the bending wavelength is considerably larger than the distance between the point connections, a continuous line junction does not necessarily represent an appropriate approximation of a point connected plate junction.

The transmission loss characterizing the conversion from bending to longitudinal waves is illustrated in Fig. 11. Up to a certain frequency, the transmission loss is slightly lower than the result for the continuous line junction. After a sudden increase, the transmission loss curve displays some intermediate minima and maxima, and tends to converge to a constant value as the frequency increases. It can be shown that at the frequency above which large deviations from the reference curve occur, the bending wave component $n = -1$ starts to interfere with longitudinal wave component $n = 0$. This phenomenon is clarified in Fig. 12, where the transmission coefficient for reflected bending waves (τ_{B1B1} , upper curves) and transmitted quasi-longitudinal waves (τ_{B1L2} , lower curves) is plotted as a function of the angle of incidence for case 3 of point spacing. The results are presented for three different frequencies in the range where

the sudden increase of transmission loss can be observed. With increasing angle of incidence, τ_{B1L2} decreases rapidly and eventually drops down to zero at the cut-off angle of the primary longitudinal wave component. At 1250 Hz, however, the transmission coefficient displays a local minimum, which corresponds to a local maximum of the reflection coefficient τ_{B1B1} near the cut-on angle of bending wave component $n = -1$. Consequently, above the frequency at which the cut-off angle of longitudinal wave component $n = 0$ equals the cut-on angle of bending wave component $n = -1$, gradually more energy is carried by reflected bending waves, resulting in the increased transmission loss in Fig. 11.

The transmission loss for incident and transmitted quasi-longitudinal waves, and the wave conversion from quasi-longitudinal to bending waves is presented in Figs. 13 and 14, respectively. In both figures, an increase of transmission loss appears above the cut-on frequency f_{LB}^{-1} . Below this frequency, the reference curve represents a good approximation of the result for transmitted longitudinal waves in Fig. 13. In Fig. 14 however, the transmission loss below f_{LB}^{-1} is slightly lower than the result for the reference case.

As a general remark, applicable to both junction geometries considered in this chapter, the difference between the transmission loss curves of two consecutive cases of point spacing approaches 3 dB with increasing frequency. Since the different cases of point spacing were obtained by changing the average number of point connections per unit length by a factor 2, the difference of 3 dB indicates that the points may be considered as acting independently.

III. EXPERIMENTAL VERIFICATION

The predictive performance of the calculation model was verified experimentally by measuring the bending wave transmission on a corner junction of two plates for various configurations of the point connections. The two identical PVC plates coupled at the junction illustrated in Fig. 15, have the following dimensions and material properties $L_x = 1.5$ m, $L_z = 1.14$ m, $h_p = 0.015$ m, Young's modulus $E = 3.7 \times 10^9$ Pa, Poisson ratio $\nu = 0.4$, and density $\rho = 1440$ kg/m³. The point connections were elaborated by inserting at the junction a number of aluminum plates with the following dimensions and material properties: $h_p = w_1 = 0.015$ m, $h_l = 0.002$ m, $E = 7.1 \times 10^{10}$ Pa, and $\nu = 0.3$. The aluminum plates were squeezed between both PVC plates using a mechanic bolt as shown in the vertically (V) and horizontally (H) projected details in Fig. 15. Each PVC plate was mounted in a separate, resiliently supported concrete frame. The plates were connected to the concrete frame at the edges normal to the junction line, while the edges opposite to the junction line remained free. To account for any vibrational energy loss at the mounted edges, the total loss factors of the plates were obtained by reverberation time measurements on the uncoupled plates, mounted in the concrete test setup. This total loss factor was introduced as an internal loss factor in the SEA calculations.

Four different junctions were considered by varying the amount of equally spaced point connections. The relative positions of the local rigid connections along the junction

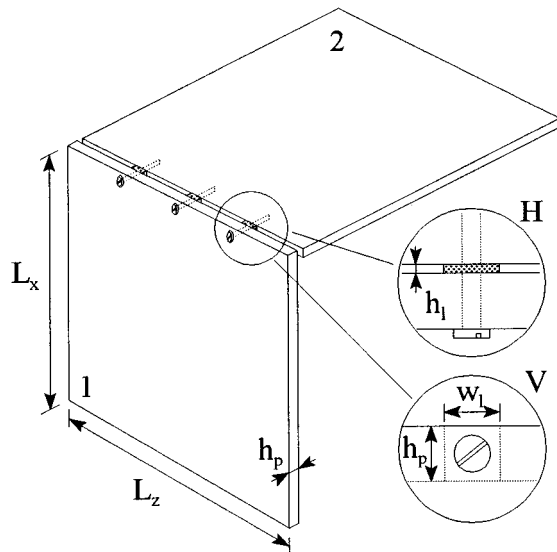


FIG. 15. Two PVC plates were coupled at distinct positions by squeezing small aluminum elements between the plates using a mechanic bolt, as indicated in the horizontally (H) and vertically (V) projected details.

line, the corresponding spatial period L and the ratio of the spatial period to the bending wavelength at 100 Hz are listed in Table II. As a reference case, additional measurements were performed on a rigid plate junction. The rigid connection between both plates was achieved by means of two aluminum corner profiles, as indicated in Fig. 16.

The bending wave transmission was evaluated by measuring the difference between the rms velocity levels of source and receiver plates (plates 1 and 2 in Fig. 15, respectively). The source plate was randomly excited using a plastic-headed hammer and the velocity level difference between the coupled plates was determined using a two channel measurement technique as proposed by Craik.¹⁵

The theoretical results were obtained by using a simple two subsystem SEA calculation scheme. For a junction of two identical plates, it can be shown that the velocity level difference L_{v12} is given as

$$L_{v12} = 10 \log \left(1 + \frac{\eta_{22}}{\eta_{12}} \right). \quad (36)$$

The coupling loss factor η_{12} is expressed in terms of the angle-averaged transmission coefficient τ_{12} .¹⁰

TABLE II. Relative positions of the centers of the local rigid connections along the junction line, the corresponding spatial period L , and the ratio of the spatial period to the bending wavelength at 100 Hz. Coupled plate edge length $L_z = 1.14$ m.

| Number of points | Relative positions z/L_z | Spatial period L [m] | $L/\lambda_{B,100 \text{ Hz}}$ |
|------------------|-----------------------------------|------------------------|--------------------------------|
| 7 | 1/8, 1/4, 3/8, 1/2, 5/8, 3/4, 7/8 | 0.1425 | 0.111 |
| 4 | 1/8, 3/8, 5/8, 7/8 | 0.285 | 0.222 |
| 3 | 1/4, 1/2, 3/4 | 0.285 | 0.222 |
| 2 | 1/4, 3/4 | 0.57 | 0.444 |

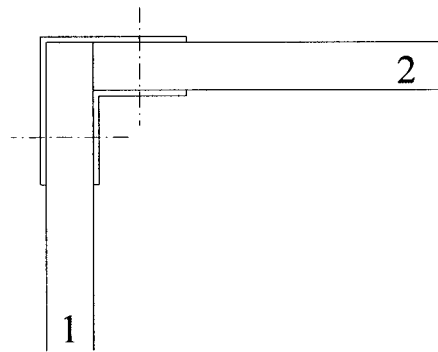


FIG. 16. A rigid junction was elaborated using two aluminum corner profiles.

$$\eta_{12} = \frac{C_{g1} l_{12} \tau_{12}}{\omega \pi S_1}. \quad (37)$$

Due to the averaging of the transmitted intensities in Eq. (27) over one spatial period of the boundary, the transmission coefficient is representative for an elementary interval of length L around each point connection. Consequently, the effective junction length l_{12} should be calculated as the number of point connections N_c times the spatial period L : $l_{12} = N_c L$. The calculations were performed at the center frequencies of each one-third octave band. The internal damping of the interlayer material was not included and the interlayer stiffness was calculated using Eqs. (34a)–(34d). The junction offset was taken as half the plate thickness.

The measured velocity level differences for the point connected plate junctions and the rigid junction are plotted as

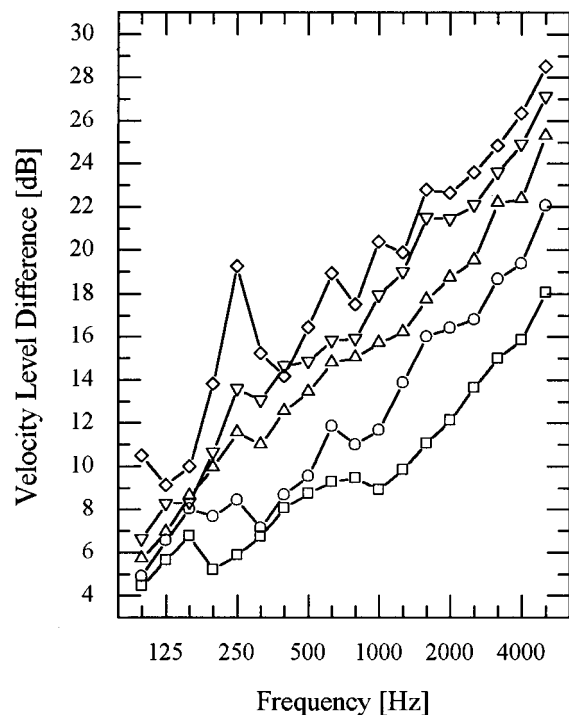


FIG. 17. Measured transverse velocity level differences: rigid junction (\square); 7 (\circ), 4 (\triangle), 3 (∇), and 2 (\diamond) point connections.

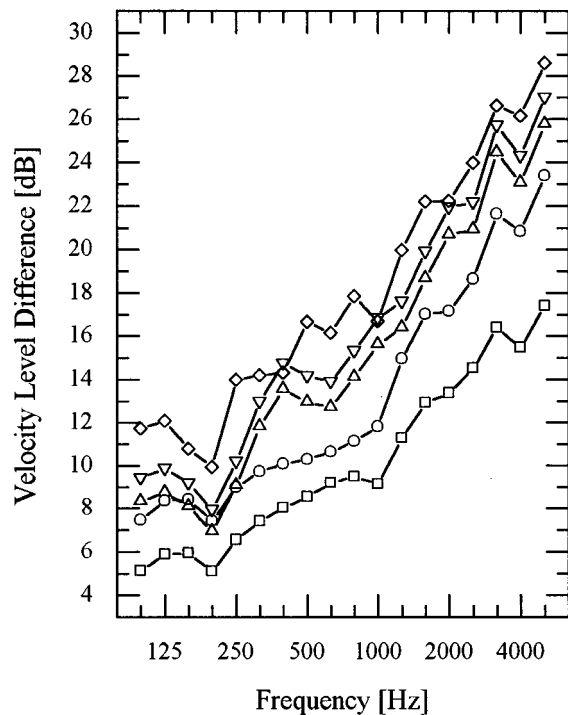


FIG. 18. Predicted transverse velocity level differences: rigid junction (□); 7 (○), 4 (△), 3 (▽), and 2 (◇) point connections.

a function of frequency in Fig. 17. As could be expected, the velocity level difference decreases with increasing number of point connections. While the largest deviations from the result for the rigid junction can be observed above 1000 Hz, the results for the point connected plate junctions approach

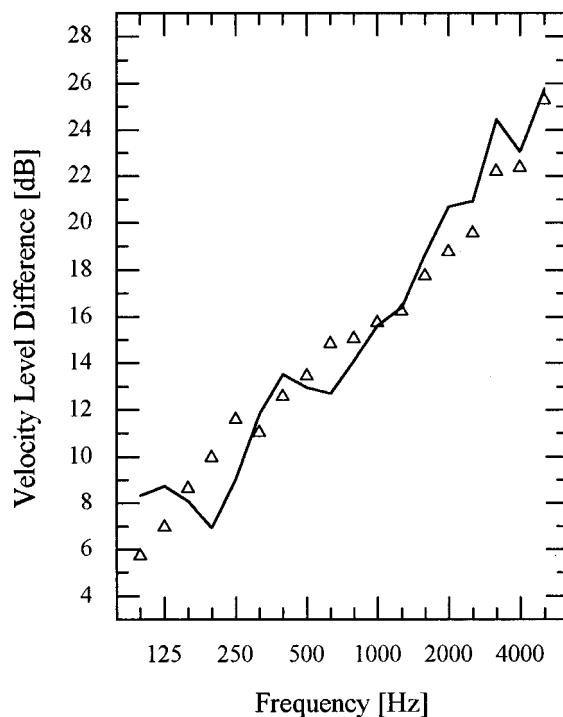


FIG. 20. Predicted (—) and measured (△) velocity level difference for the junction with 4 point connections.

the reference curve as the frequency decreases. However, the velocity level difference for the case with the highest density of point connections exceeds the result for the rigid junction in the full frequency range, confirming the findings of the previous chapter. By comparing the measurement data in

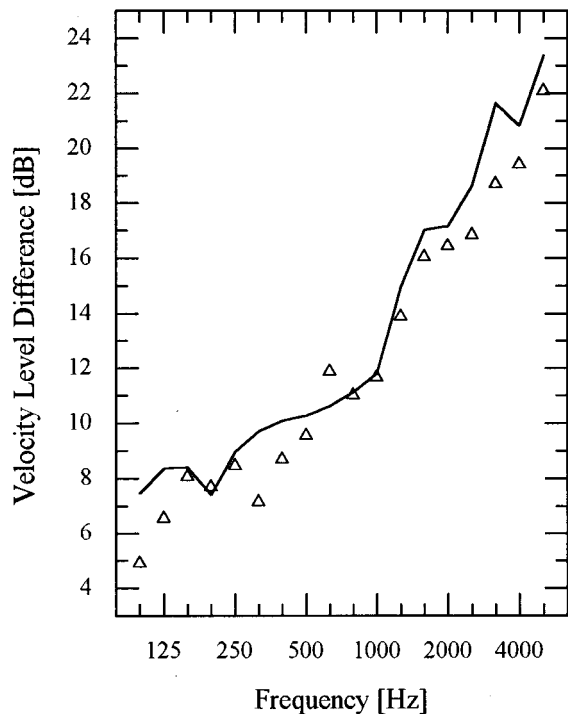


FIG. 19. Predicted (—) and measured (△) velocity level difference for the junction with 7 point connections.

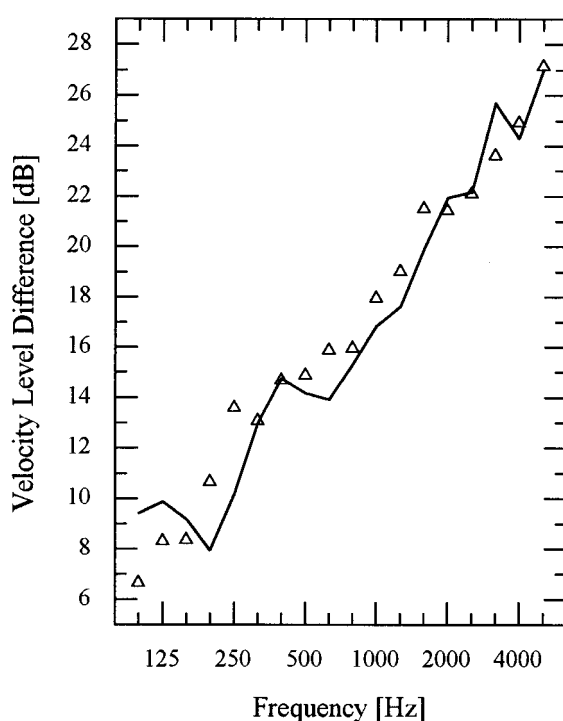


FIG. 21. Predicted (—) and measured (△) velocity level difference for the junction with 3 point connections.

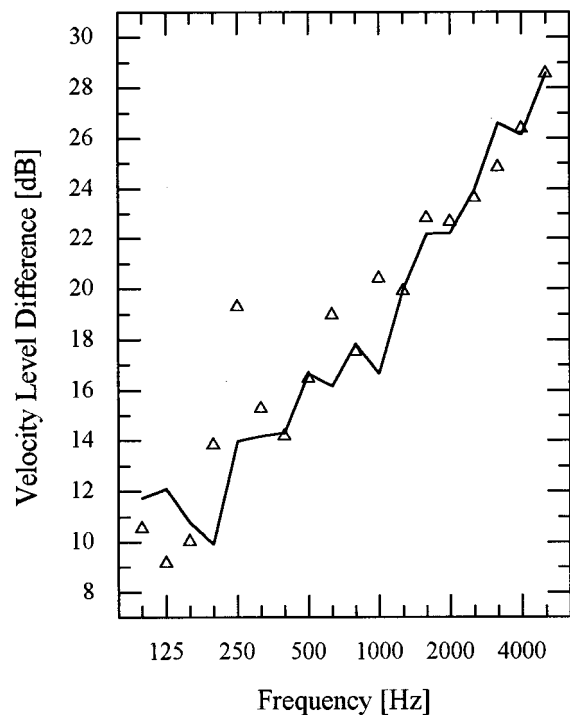


FIG. 22. Predicted (—) and measured (Δ) velocity level difference for the junction with 2 point connections.

Fig. 17 to the theoretical results in Fig. 18, it is clearly demonstrated that global trends are reliably predicted by the calculation model. At low frequencies, however, the velocity level difference for the point connected plate junctions tend to be overestimated by the SEA calculations. Further, a more detailed comparison in Figs. 19–22 shows that the accuracy of the SEA predictions increases with increasing frequency. The largest deviation (5.3 dB) can be observed in the one-third octave band of 250 Hz for the case of two point connections (Fig. 22). In general, however, the comparison shows a good agreement between measurement and calculation since the frequency averaged prediction error is less than 1.5 dB for all cases.

Apart from the point spacing L , the width w_1 of the local rigid connections represents an important parameter to be dealt with when designing a junction of plates coupled by local rigid connections. The influence of the width was assessed experimentally and theoretically for the junction with four point connections. In the measurement setup, two additional aluminum plates were positioned on both sides of the existing connections to triple the contact length from 15 mm to 45 mm. The measured and predicted velocity level difference between source and receiver plate are plotted in Fig. 23. Again, a good agreement can be observed in the full frequency range. The lower part of Fig. 23 shows the difference between the results for the junctions with 15 mm (Fig. 20) and 45 mm wide connections, $L_{v12,15\text{ mm}} - L_{v12,45\text{ mm}}$, demonstrating the fact that structure-borne sound increases with increasing width w_1 . Both the measured and predicted differences oscillate around a frequency averaged value of 2 dB.

If the local rigid connections were assumed to act as small line junctions, one could consider to estimate the cou-

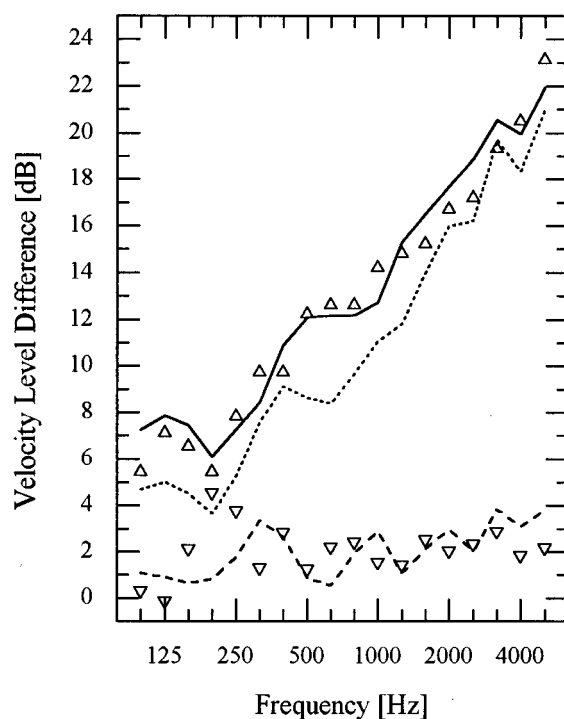


FIG. 23. Predicted (—) and measured (Δ) velocity level difference for the junction with 4, 45 mm wide point connections. Difference between the results for the 15 mm and 45 mm wide connections $L_{v12,15\text{ mm}} - L_{v12,45\text{ mm}}$: predicted (---) and measured (∇). Predicted velocity level difference for the 45 mm wide connections using $\eta_{12,45\text{ mm}} = 3 \eta_{12,15\text{ mm}}$ (.....).

pling loss for the 45 mm wide connections as a function of the coupling loss factor for the 15 mm wide connections: $\eta_{12,45\text{ mm}} = 3 \eta_{12,15\text{ mm}}$. After substitution of this coupling loss factor into Eq. (36), a velocity level difference is obtained as illustrated in Fig. 23, where a clear overestimation of the structure-borne sound transmission can be noticed. Only in the highest one-third octave bands an acceptable accuracy is achieved. This observation suggests that when the bending wavelength is larger than the width of the coupling elements, a more advanced calculation model is required to estimate the energy flow through the junction.

IV. CONCLUSIONS

The structure-borne sound transmission at junctions of point connected plates has been studied theoretically and experimentally. A calculation model has been presented for the general case of a junction of semi-infinite plates coupled by an interlayer with periodically varying elastic properties. A Fourier decomposition technique was used to describe the plate response, and expressions for the boundary conditions were derived using an integral expression formulation. The influence of the distance between the point connections was investigated numerically on a number of plate junctions for different cases of point spacing. In general, it is found that structure-borne sound transmission increases with increasing ratio of the wavelength to the distance between the connections. Although most of the phenomena characterizing the energy flow at point connected plate junctions are related to

the interaction of different wave components, the interpretation of the results are difficult for junctions involving wave conversion. One remarkable observation is that for only relatively large ratios of the wavelength to the point spacing, does a continuous line junction represent a good approximation of a point connected corner junction. Further, the SEA assumption of independently acting point connections is justified only above the cut-on frequencies of all secondary wave components $n = -1$. Finally, measurements were carried out on several variants of a corner junction of PVC plates and good agreement was achieved between experimental and numerical results.

ACKNOWLEDGMENT

This research was funded by a grant from the Flemish Institute for the promotion of Scientific Technological Research in Industry (IWT).

- ¹P. Mees and G. Vermeir, "Structure-borne sound transmission at elastically connected plates," *J. Sound Vib.* **166**, 55–76 (1993).
- ²R. J. M. Craik and A. G. Osipov, "Structural isolation of walls using elastic interlayers," *Appl. Acoust.* **46**, 233–249 (1995).
- ³A. G. Osipov and G. Vermeir, "Sound transmission in buildings with elastic layers at joints," *Appl. Acoust.* **49**, 141–162 (1996).

- ⁴L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound* (Springer-Verlag, Berlin, 1988).
- ⁵T. Kihlman, *Transmission of Structure-Borne Sound Through Buildings*, Report 9 (National Swedish Institute for Building Research, Stockholm, 1967).
- ⁶W. Wöhle, Th. Beckmann, and H. Schreckenbach, "Coupling loss factors for statistical energy analysis of sound transmission at rectangular structural slab joints, Part I," *J. Sound Vib.* **77**, 323–334 (1981).
- ⁷W. Wöhle, Th. Beckmann, and H. Schreckenbach, "Coupling loss factors for statistical energy analysis of sound transmission at rectangular structural slab joints, Part II," *J. Sound Vib.* **77**, 335–344 (1981).
- ⁸P. G. Craven and B. M. Gibbs, "Sound transmission and mode coupling at junctions of thin plates, Part I: Representation of the problem," *J. Sound Vib.* **77**, 417–427 (1981).
- ⁹B. M. Gibbs and P. G. Craven, "Sound transmission and mode coupling at junctions of thin plates, Part II: Parametric survey," *J. Sound Vib.* **77**, 429–435 (1981).
- ¹⁰R. H. Lyon and R. G. DeJong, *Theory and Application of Statistical Energy Analysis* (Butterworth-Heinemann, Boston, 1995), 2nd ed., p. 182.
- ¹¹J. E. Manning, "Formulation of SEA parameters using mobility functions," *Philos. Trans. R. Soc. London, Ser. A* **346**, 477–488 (1994).
- ¹²R. J. M. Craik, *Sound Transmission through Buildings using Statistical Energy Analysis* (Gower, Hampshire, 1996), p. 120.
- ¹³R. S. Langley and K. H. Heron, "Elastic wave transmission through plate/beam junctions," *J. Sound Vib.* **143**, 241–253 (1990).
- ¹⁴Y. Ando, *Concert Hall Acoustics* (Springer-Verlag, Berlin, 1985), p. 12.
- ¹⁵R. J. M. Craik, "The measurement of structure-borne sound transmission using impulsive sources," *Appl. Acoust.* **15**, 355–361 (1982).

Frequency response of a two-dimensional trusslike periodic panel

Michael El-Raheb

The Dow Chemical Company, Midland, Michigan 48674

(Received 29 August 1996; accepted for publication 7 January 1997)

The frequency response of a two-dimensional trusslike periodic panel is studied adopting the transfer matrix of a repeated cell. Intracell resonance occurs at relatively low frequencies, causing high modal density of the panel. Effects of panel curvature, sag of cell diagonal member, mass added by a liquid filling the cell cavity, and damping by constrained viscoelastic layers are studied parametrically. © 1997 Acoustical Society of America. [S0001-4966(97)01705-0]

PACS numbers: 43.40.Dx [CBB]

INTRODUCTION

Lightweight transparent polymer panels are attractive components in plans to lessen noise from highways and railways, especially in densely populated cities. One major advantage of transparent panels over opaque concrete walls is architectural esthetics and illumination by sunlight in surface tunnels. The light weight of a panel reduces expense and labor to assemble over long paths. To achieve structural integrity while maintaining light weight, a trusslike construction was suggested that includes repeated thin walled rectangular cells reinforced by a diagonal member.

Frequency response of periodic structures has been studied extensively in the literature.¹⁻⁸ Floquet theory was applied to beams, plates, shells, and layered composites to define propagation and attenuation zones in the frequency domain. The mathematical procedure determines propagation constants from the eigenvalues of the cell transfer matrix. As the structure of a cell becomes more complicated from extension and flexure of several members, the order of cell transfer matrix rises and so does the difficulty in identifying propagation constants related to specific motions of the cell.⁹⁻¹² An example facing this difficulty is the trusslike panel treated in this work. In two dimensions, the cell transfer matrix is of order 12 yielding six distinct complex pairs of propagation constants. Also, for panels constructed from a viscoelastic polymerlike polycarbonate with high viscous dissipation, identification of propagation constants becomes even more difficult.

The purpose of this work is to determine the frequency response of a 2-D trusslike periodic panel and evaluate parametrically the sensitivity of response to panel curvature, sag of diagonal member, and mass of fluid filling the cavity of cells. It was found early in this development that intracell resonances start at relatively low frequencies due to thinness of the members. This in turn substantially raises modal density. A solution by finite elements would require 40 elements per cell which, for an 80 cell panel, totals 9600 degrees of freedom. In the frequency domain, a direct solution at intervals of 10 Hz over a range of 8 kHz becomes computationally intensive and the extension to three dimensions is prohibitive.

In view of these considerations, an analytical method based on a 2-D transfer matrix of a cell is adopted. The

transfer matrix relates state vector of forces and displacements at two corners of a cell to that at the other two corners. Applying continuity of state vector at interfaces of cells produces a set of simultaneous equations in tridiagonal form. Sag of a member is caused by gravity during extrusion of the panel while the polymer is hot. The geometry of the sagging member is approximated by the arc of a circle for which a transfer matrix is derived. The consequence of eliminating the diagonal member is also discussed. Finally, damping treatment by thin constrained viscoelastic layers bonded to the panel^{13,14} is examined as a means to reduce transmission mobility.

I. ANALYSIS

Figure 1(a) shows a 3-D view of the trusslike panel with periodicity along the global x axis. It also shows the 2-D periodic strip simulated in the present analysis. A cell of the 2-D periodic truss is shown in Fig. 1(b). Transfer matrix of a cell is found in terms of transfer matrices of the constituent members. Let $\{\mathbf{f}, \mathbf{g}\}$ be the state vector at one end of a member

$$\{\mathbf{f}, \mathbf{g}\}^T = \{f_x, f_y, m_z, u_x, u_y, \theta_z\}^T, \quad (1)$$

where (x, y) are local axial and transverse coordinates of a member. The member transfer matrix \mathbf{T}_i is defined by

$$\{\mathbf{f}, \mathbf{g}\}_{i2}^T = \mathbf{T}_i \{\mathbf{f}, \mathbf{g}\}_{i1}^T, \quad (2)$$

where subscripts 1 and 2 refer to the two ends of the i th member. \mathbf{T}_i is split into submatrices

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{t}_{i1} & \mathbf{t}_{i2} \\ \mathbf{t}_{i3} & \mathbf{t}_{i4} \end{bmatrix}. \quad (3)$$

Let \mathbf{a}_i be a coordinate transformation matrix from local member coordinates (x, y) to global panel coordinates (X, Y) . At each of the four corners of the cell, enforcing equilibrium of \mathbf{f} and replacing $\{\mathbf{f}, \mathbf{g}\}_{i2}$ with $\{\mathbf{f}, \mathbf{g}\}_{i1}$ using (2) yields

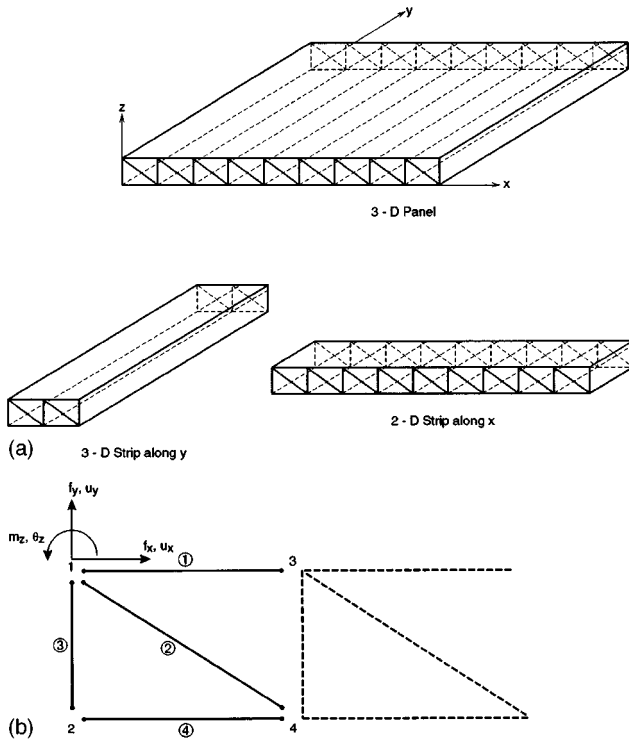


FIG. 1. (a) Panel with cell periodicity along one axis. (b) Periodic trusslike cell.

$$\begin{aligned} \mathbf{A}\mathbf{f}_c + \bar{\mathbf{A}}\mathbf{g}_c &= \mathbf{f}^*, \\ \mathbf{f}_c &= \{f_{11}, f_{21}, f_{31}, f_{41}\}^T, \quad \mathbf{g}_c = \{g_{11}, g_{21}, g_{31}, g_{41}\}^T, \\ \mathbf{f}^* &= \{f_1^*, f_2^*, f_3^*, f_4^*\}, \end{aligned} \quad (4)$$

$\{\mathbf{f}_c, \mathbf{g}_c\}$ are local state vectors at end 1 of all members, $\{\mathbf{f}^*, \mathbf{g}^*\}$ are global state vectors at the four corners of the cell, and

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{a}_3 \mathbf{t}_{31} & \mathbf{a}_4 \\ -\mathbf{a}_1 \mathbf{t}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{a}_2 \mathbf{t}_{21} & \mathbf{0} & -\mathbf{a}_4 \mathbf{t}_{41} \end{bmatrix}, \quad (4a)$$

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{a}_3 \mathbf{t}_{32} & \mathbf{0} \\ -\mathbf{a}_1 \mathbf{t}_{12} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{a}_2 \mathbf{t}_{22} & \mathbf{0} & -\mathbf{a}_4 \mathbf{t}_{42} \end{bmatrix}. \quad (4b)$$

Applying continuity of \mathbf{g}_c at each corner yields

$$\begin{aligned} \mathbf{g}_c &= \mathbf{D}\mathbf{g}^* \\ \mathbf{D} &= \begin{bmatrix} \mathbf{a}_1^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{a}_2^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{a}_3^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_4^{-1} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \end{aligned} \quad (5)$$

Eliminating \mathbf{g}_c from (4) and (5) gives

$$\mathbf{A}\mathbf{f}_c + \mathbf{B}\mathbf{g}^* = \mathbf{f}^*, \quad \mathbf{B} = \bar{\mathbf{A}}\mathbf{D}. \quad (6)$$

Expressing \mathbf{f}_c in terms of \mathbf{g}^* by continuity of \mathbf{g}_c at corners 2, 3, and 4 while invoking Eq. (2) gives

$$\begin{aligned} \mathbf{f}_c &= \mathbf{C}\mathbf{g}^*, \\ \mathbf{C} &= \begin{bmatrix} -\mathbf{t}_{13}^{-1} \mathbf{t}_{14} \mathbf{a}_1^{-1} & \mathbf{0} & \mathbf{t}_{13}^{-1} \mathbf{a}_1^{-1} & \mathbf{0} \\ -\mathbf{t}_{23}^{-1} \mathbf{t}_{24} \mathbf{a}_2^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{t}_{23}^{-1} \mathbf{a}_2^{-1} \\ -\mathbf{t}_{33}^{-1} \mathbf{t}_{34} \mathbf{a}_3^{-1} & \mathbf{t}_{33}^{-1} \mathbf{a}_3^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{t}_{43}^{-1} \mathbf{t}_{44} \mathbf{a}_4^{-1} & \mathbf{0} & \mathbf{t}_{43}^{-1} \mathbf{a}_4^{-1} \end{bmatrix}. \end{aligned}$$

Eliminating \mathbf{f}_c from (6) and (7) yields

$$\mathbf{f}^* = [\mathbf{AC} + \mathbf{B}]\mathbf{g}^* = \mathbf{Z}_c \mathbf{g}^*. \quad (8)$$

Stiffness matrix \mathbf{Z}_c relates global force vectors \mathbf{f}^* at the four corners of a cell to global displacement vectors \mathbf{g}^* at those corners. Here, \mathbf{Z}_c is converted to a transfer matrix by rewriting ($\mathbf{f}^*, \mathbf{g}^*$) as

$$\mathbf{f}^* = \{\mathbf{f}_L^*, \mathbf{f}_R^*\}, \quad \mathbf{g}^* = \{\mathbf{g}_L^*, \mathbf{g}_R^*\}, \quad \mathbf{Z}_c = \begin{bmatrix} \mathbf{Z}_{LL} & \mathbf{Z}_{LR} \\ \mathbf{Z}_{RL} & \mathbf{Z}_{RR} \end{bmatrix}, \quad (9)$$

where subscripts L and R refer to left and right corners of a cell. Solving for $\{\mathbf{f}_R^*, \mathbf{g}_R^*\}$ in terms of $\{\mathbf{f}_L^*, \mathbf{g}_L^*\}$ produces

$$\{\mathbf{f}_R^*, \mathbf{g}_R^*\} = \mathbf{T}_c \{\mathbf{f}_L^*, \mathbf{g}_L^*\}, \quad (10)$$

$$\mathbf{T}_c = \begin{bmatrix} \mathbf{Z}_{RR} \mathbf{Z}_{LR}^{-1} & \mathbf{Z}_{RL} - \mathbf{Z}_{RR} \mathbf{Z}_{LR}^{-1} \mathbf{Z}_{LL} \\ \mathbf{Z}_{LR}^{-1} & -\mathbf{Z}_{LR}^{-1} \mathbf{Z}_{LL} \end{bmatrix},$$

where \mathbf{T}_c is cell transfer matrix in global panel coordinates. Transfer matrix of a member \mathbf{T}_i is derived in Appendix A. \mathbf{T}_c of the rightmost cell differs from all other \mathbf{T}_c because it includes member (3–4). Details of derivations for this case follow the same procedure as that of any other cell and are omitted for shortness.

Continuity of $\{\mathbf{f}_R^*, \mathbf{g}_R^*\}_n$ of cell (n) and $\{\mathbf{f}_L^*, \mathbf{g}_L^*\}_{n+1}$ of cell ($n+1$) with use made of (10) determines simultaneous equations in tridiagonal form:

$$\begin{bmatrix} \mathbf{I} & \mathbf{z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{c1} & \mathbf{T}_{c2} & \mathbf{I} & \mathbf{z}_2 \\ \mathbf{T}_{c3} & \mathbf{T}_{c4} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{c1} & \mathbf{T}_{c2} & \mathbf{I} & \mathbf{z}_3 \\ & & \mathbf{T}_{c3} & \mathbf{T}_{c4} & \mathbf{0} & -\mathbf{I} & \dots \\ & & \mathbf{0} & \mathbf{0} & \mathbf{T}_{c1} & \mathbf{T}_{c2} & \dots \\ & & & & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} \mathbf{f}_{L1} \\ \mathbf{g}_{L1} \\ \mathbf{f}_{L2} \\ \mathbf{g}_{L2} \\ \mathbf{f}_{L3} \\ \mathbf{g}_{L3} \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{f}_{01} \\ \mathbf{f}_{02} \\ \mathbf{0} \\ \mathbf{f}_{03} \\ \mathbf{0} \\ \mathbf{f}_{04} \\ \vdots \end{bmatrix}, \quad (11)$$

where \mathbf{f}_{0n} is external force vector at nodes 1 and 2 of the n th cell, \mathbf{I} is the unit matrix, \mathbf{z}_i is a stiffness matrix at corners 1 and 2 of the i th interface and \mathbf{T}_{c_j} are constituent submatrices of \mathbf{T}_c similar to (3). When written in matrix form, Eq. (11) becomes

$$\mathcal{L}\mathcal{S} = \mathcal{F}_0, \quad (12)$$

where \mathcal{L} is the assembled global tridiagonal matrix in (11) made of cell transfer matrices \mathbf{T}_c and boundary conditions, \mathcal{S} is the assembled state vector at all interfaces, and \mathcal{F}_0 is the assembled external force vector. Boundary constraints at the two ends of the truss are considered by including linear

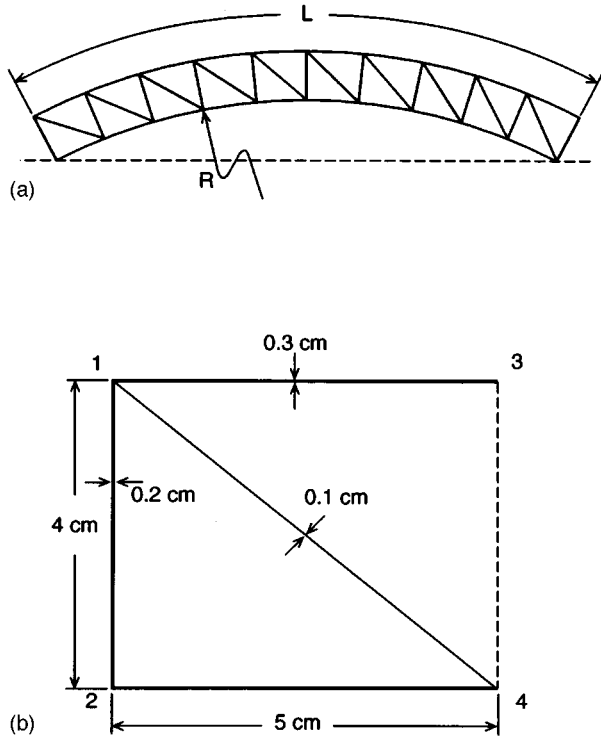


FIG. 2. (a) Typical panel. (b) Cell geometry.

springs along local cell axes x and y at corner points 1 and 2 of cell 1, and at corner points 3 and 4 of cell N , where N is the number of cells in the truss. These springs appear in the stiffness matrices \mathbf{z}_1 and \mathbf{z}_{n+1} of \mathcal{C} . Mass added by fluid filling the cavity of cells, springs linking a corner point of a cell to a rigid boundary, and damping by a constrained viscoelastic layer are constituents of \mathbf{z}_i as derived in Appendix B.

II. RESULTS

A typical panel is shown in Fig. 2(a) with cell dimensions given in Fig. 2(b). Material properties of polycarbonate are

$$E = 2.3 \text{ GPa}, \quad \rho = 1200 \text{ kg/m}^3, \quad \nu = 0.40.$$

A constant viscoelastic damping is assumed in the form of an imaginary part of Young's modulus $E = E_0(1 + i\zeta)$. A value of $\zeta = 0.02$ is considered for all the examples to follow. To simulate simply supported ends, spring stiffnesses at the corners of end cells are assumed to be $k_x = 10^2 \text{ dyn/cm}$ and $k_y = 10^9 \text{ dyn/cm}$. The panel is excited by single-frequency point forces f_{Y_0} of magnitude 1 dyn acting along Y at point 1 of each interface to simulate a uniform pressure distribution $p_0 = f_{Y_0}/w_c$ of 0.2 dyn/cm^2 where w_c is cell width. A measure of response is root-mean square of velocity v_{rms} summed over all excited points along Y

$$v_{\text{rms}} = \left(\sum_{n=1}^N (\omega u_Y)^2_n \right)^{1/2}, \quad (13a)$$

where ω is radian frequency of excitation and u_Y is displacement along Y . v_{rms} is chosen to measure response since it is

related to energy in motion of the panel along Y and sound-pressure level radiated in the medium. An intensity I_{rms} in dB is defined as

$$I_{\text{rms}} = 20 \log_{10}(v_{\text{rms}}/v_0), \quad (13b)$$

where $v_0 = 5 \times 10^{-6} \text{ cm/s}$ is a reference velocity. Approximating $u_Y(X)$ by

$$u_Y(X) = u_{Y_{\text{max}}} \sin(m\pi X/L) \quad (13c)$$

then substituting in (13a) noting that

$$\sum_{j=1}^N \sin^2(m\pi j w_c/L) = N/2 \quad (13d)$$

yields

$$v_{\text{rms}} \approx \omega u_{Y_{\text{max}}} \sqrt{N/2} \Rightarrow I_{\text{rms}} \approx 20 \log_{10}(\omega u_{Y_{\text{max}}}/v_0) + 10 \log_{10}(N/2). \quad (13e)$$

Relation (13e) reveals that I_{rms} varies smoothly with number of cells N .

Deformed shapes of a circular panel with length $L = 2 \text{ m}$ and radius $R = 2 \text{ m}$ are shown in Fig. 3 for $100 \leq \Omega \leq 600 \text{ Hz}$. For each frequency, enlarged deformed cells at $1/4L$, $1/2L$, and $3/4L$ also appear. Up to 100 Hz, deformation is caused by axial strain of cell members with negligible flexure. Above 300 Hz, flexure of the diagonal member is noticeable. This applies to the other members above 600 Hz. Figure 4 compares frequency response of a flat panel with $L = 2 \text{ m}$ for three cases: (a) with flexural inertia, (b) without flexural inertia, (c) cell members are hinged at both ends, allowing axial loads only. The transfer matrix of the "true truss" member in case (c) is developed in Appendix C. Omitting flexural inertia suppresses intracell flexural resonances, restricting members to resonate axially. The responses differ substantially in two ways. Flexural inertia reduces response since a portion of the input energy produces flexure, leaving a smaller portion for extension. Also, flexural inertia changes all peaks and valleys of I_{rms} response. This demonstrates that flexural inertia is essential in simulating panel dynamics.

Certain features of frequency response shown as the solid line in Fig. 4 are typical of periodic structures. Response is divided into frequency zones. The first zone has the following features:

- Frequencies are below the fundamental intracell resonances ($\Omega < 300 \text{ Hz}$).
- Panel response modes can be approximated by those of an equivalent homogeneous panel with appropriate choice of material properties and thickness [see Fig. 3(a)–(c)].^{16,17}
- Resonances of the panel are weakly affected by structural damping as evidenced by the sharpness of peaks.
- Average I_{rms} achieves a maximum $(I_{\text{rms}})_{\text{max}1}$ at the fundamental resonance of the panel, then diminishes with frequency reaching its first minimum $(I_{\text{rms}})_{\text{min}1}$ near 1500 Hz when the highest fundamental member resonance is crossed. In the range $900 \text{ Hz} < \Omega < 1600 \text{ Hz}$, response modes of the truss lose their waviness.

The second zone has the following features:

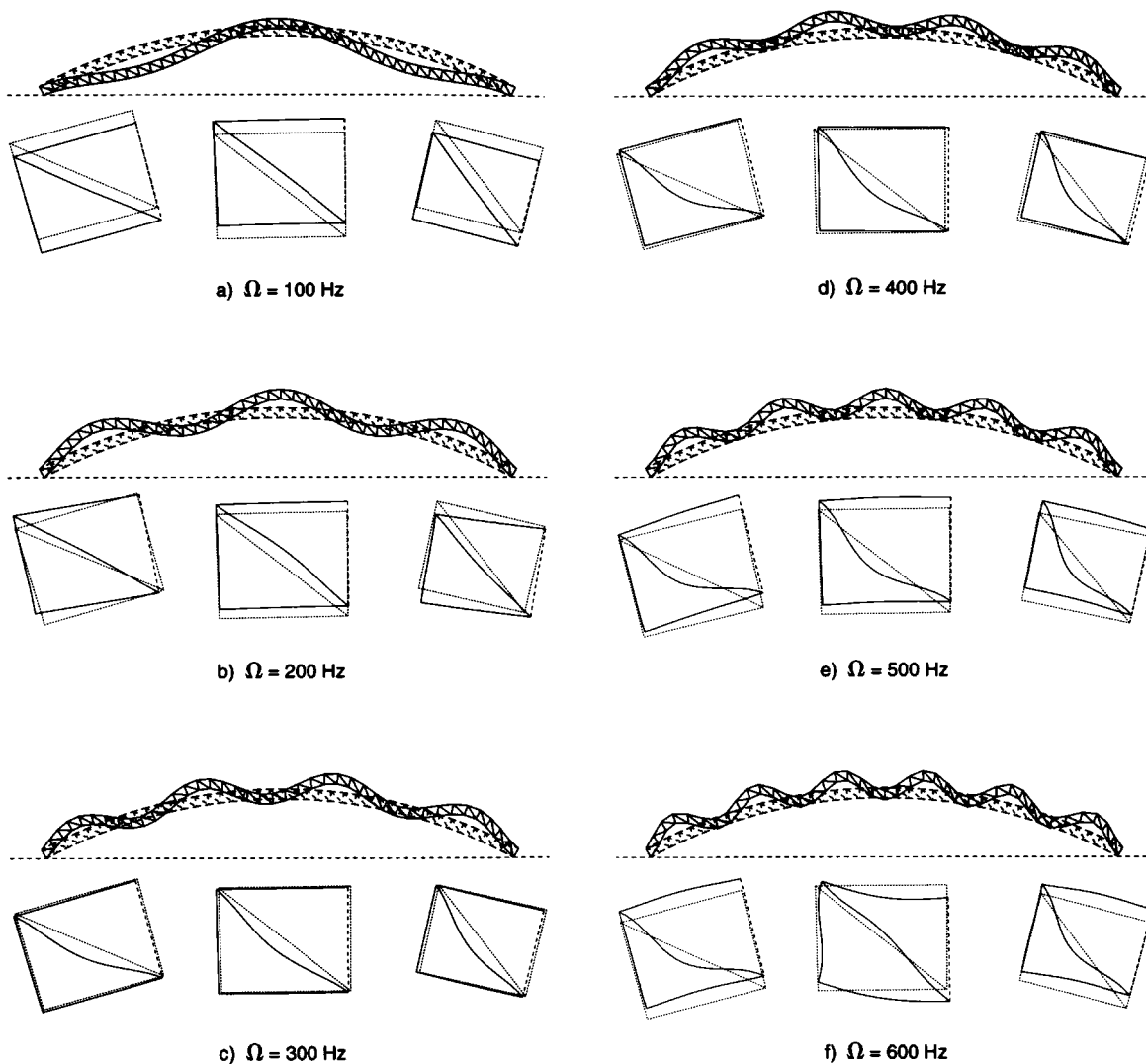


FIG. 3. Response modes of panel and intracell motion, $L=2$ m, $R=2$ m, $40 \times (5 \times 4$ cm cell).

- (a) Equivalent homogenization fails because response is dominated by flexure of cell members.
- (b) Response is not sensitive to panel length and curvature, but depends only on cell geometry.
- (c) I_{rms} rises again along a line excluding sharp resonant peaks because of damping and reaches a second maximum $(I_{\text{rms}})_{\text{max}2}$ near 2400 Hz. In the range 1800 Hz $< \Omega < 2700$ Hz, response modes regain their waviness.

This repeated pattern of maxima and minima in I_{rms} accompanied by waviness or the lack of it in response modes is reminiscent of propagation and attenuation zones found in simpler structures.¹⁻⁸

The effect on response of panel curvature is shown in Fig. 5 for a panel with $L=2$ m and $R=\infty$, and 1 m. Panel curvature has a small effect on response below 200 Hz, prior to intracell resonances. Above 500 Hz, response is insensitive to curvature. An explanation is that response is most affected by curvature when wavelength λ is comparable to R and the angle subtended by λ is largest. As λ diminishes, the angle is reduced, ultimately approaching the case of a flat panel.

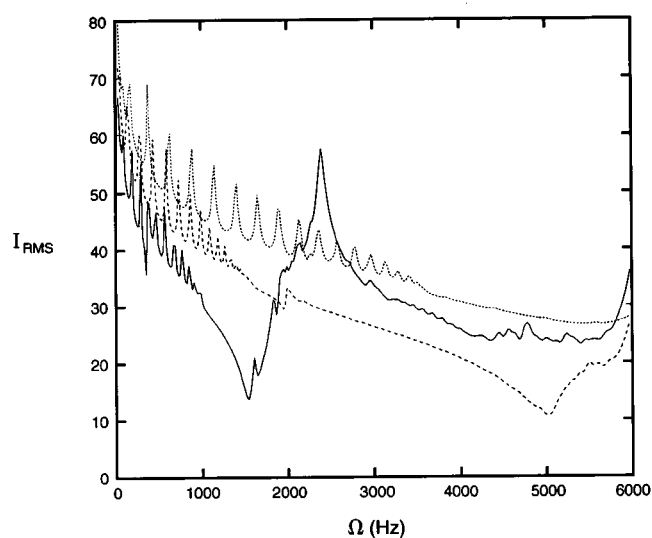


FIG. 4. Frequency response of I_{rms} for three models: (1) solid line: flexural inertia included; (2) short dashed line: members bend but flexural inertia suppressed; (3) dashed line: members treated as rods hinged at both ends. $L=2$ m, $R=\infty$, $\zeta=0.02$, $40 \times (5 \times 4$ cm cell).

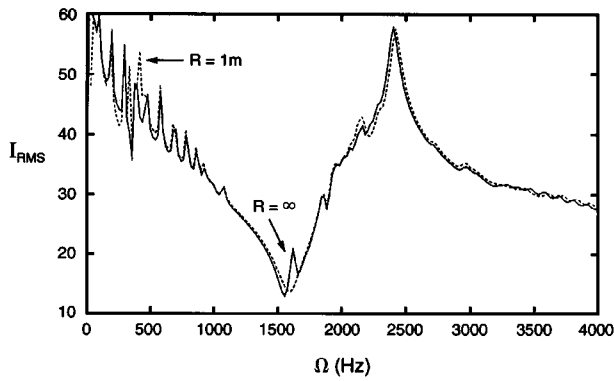


FIG. 5. Effect of panel curvature on frequency response, solid line: $R = \infty$, dashed line: $R = 1$ m. $L = 2$ m, $40 \times (5 \times 4)$ cm cells.

The effect on static deformation of diagonal sag is evaluated before examining its dynamics. Sag “ e ” is measured by maximum distance of a circular arc from its chord. Assume flat panels of varying length keeping cell and sag of the diagonal member the same. The ratio of maximum panel deformation without and with sag δ_0/δ_e is a measure of stiffness reduction from diagonal sag, termed stiffness efficiency S_e . When the ratio approaches unity, stiffness of the panel with perfect diagonals is recovered. Figure 6(a) shows how $S_e = \delta_0/\delta_e$ varies with panel L/w_c with e as parameter where w_c is cell width. When $L/w_c = 20$, a 0.3-cm sag reduces stiffness to 63% that of the perfect panel. Also shown is the limiting case of a panel without diagonals. For a constant e , S_e rises with L/w_c and approaches unity for large L/w_c . In a cell, the contribution to stiffness of its diagonal member is proportional to axial strain which increases with shear force at an interface of cells. Since shear force is proportional to $(w_c/L)^3$, then the effect of the diagonal on panel static stiffness is measured by $S_\infty \equiv \delta_0/\delta_\infty$ where δ_∞ is maximum displacement of the panel without diagonals. It follows that

$$S_\infty \equiv \frac{\delta_0}{\delta_\infty} = \frac{K_f}{K_f + (w_c/L)^3 K_{d0}}, \quad (14a)$$

where K_f is contribution to cell flexural stiffness from the flanges caused by rigidity of cell corners, and

$$K_{d0} = \pi^3 (h_d/l_d) \sin^2 \psi \quad (14b)$$

is a geometric factor to cell stiffness from the diagonal, where $[h_d, l_d]$ is diagonal thickness and length and ψ is the angle between diagonal and flange. Indeed $S_\infty \rightarrow 0$ as $(L/w_c) \rightarrow 0$ and $S_\infty \rightarrow 1$ as $(L/w_c) \rightarrow \infty$ consistent with Fig. 6(a). This behavior was verified experimentally.

The effect on frequency response of diagonal sag is shown in Fig. 6(b) for a flat panel with $L = 2$ m, $e = 0$ and 0.5 cm. Sag of the diagonal reduces I_{rms} when $\Omega < 1000$ Hz, keeping the position of the first minimum of I_{rms} the same. Sag shifts the position $(I_{rms})_{\max 2}$ to 2200 Hz, which in turn reduces I_{rms} above that frequency. An explanation is that curvature of a circular member lowers its axial stiffness consistent with the shift in peak in Fig. 6(b).

Similar to the static case, when e is large response approaches that of a panel without diagonals. In this case, stiff-

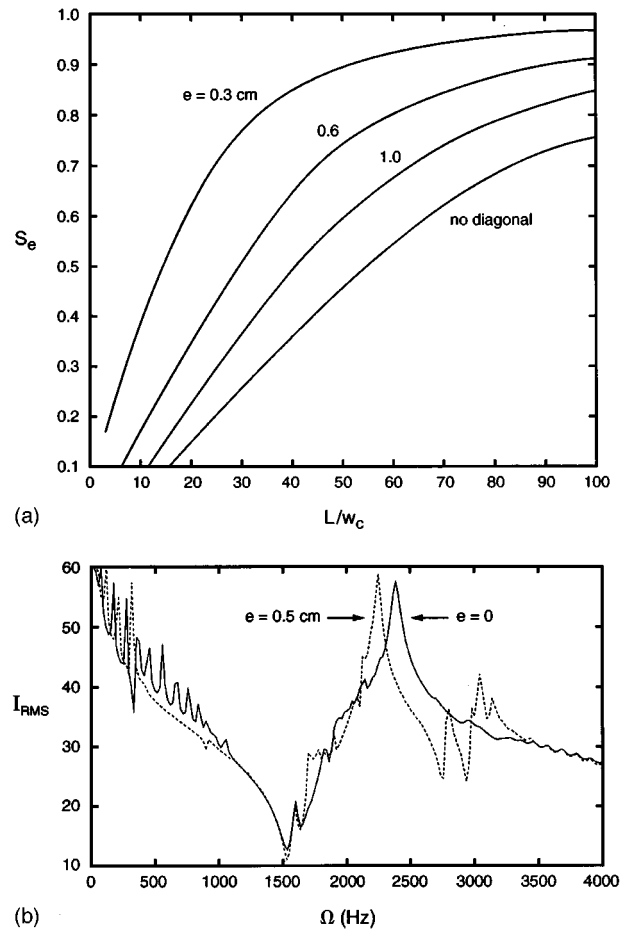


FIG. 6. (a) Variation of static stiffness efficiency, S_e with panel L/w_c and e as parameter. (b) Effect of diagonal sag on frequency response, solid line: $e = 0$, dashed line: $e = 0.5$ cm. $L = 2$ m, $40 \times (5 \times 4)$ cm cells, $R = \infty$.

ness comes primarily from bending of cell members and rigidity of its corners. Figure 7 compares response of the panel with and without diagonals. The effect of the diagonals on response depends on frequency. As expected, elimination of the diagonals raises I_{rms} below the fundamental resonance of the panel due to loss of axial stiffness [see Fig. 7(a)]. When the fundamental panel resonance is crossed and below 1000 Hz, I_{rms} is reduced. In the panel without diagonals, most of the input energy is transformed to flexure of cell members which reduces translation of its corners. In the panel with diagonals, axial stiffness of the diagonal transmitted through the path of corners diminishes because short flexural wavelength of the members isolates this path. The largest increase in I_{rms} occurs near $(I_{rms})_{\min 1}$ at 1500 Hz. For $\Omega > 2500$, extensional oscillations dominate response, which explains the higher values of I_{rms} for the panel without diagonals.

Mass of water in each cell is considered to be a lumped rigid mass applied at corners 1 and 2 of each cell and included in the stiffness matrix \mathbf{z}_i [see Appendix B, Eq. (B1)]. The effect on response of mass added by water filling the cavity of cells is demonstrated in Fig. 8(a) and (b) for panels with and without diagonals, respectively. At frequencies below $(I_{rms})_{\min 1}$, added mass reduces response by 12 dB, because fluid mass is three times that of the structure and mo-

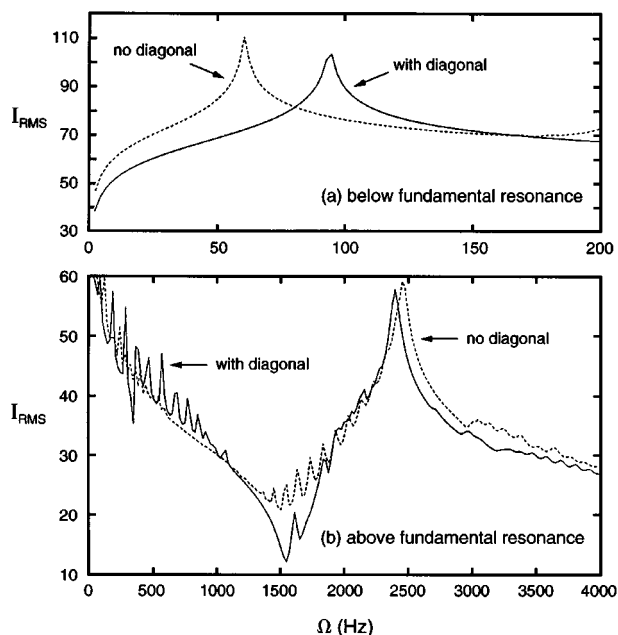


FIG. 7. Effect of diagonal member on frequency response, solid line: with diagonal, dashed line: no diagonal. $L = 2$ m, $40 \times (5 \times 4)$ cm cells, $R = \infty$.

tion of cells is primarily translational. Response close to $(I_{rms})_{min1}$ is no longer controlled by flexural inertia of the cell but by translational inertia of water. This in turn raises translation at the cell corners increasing response. Beyond

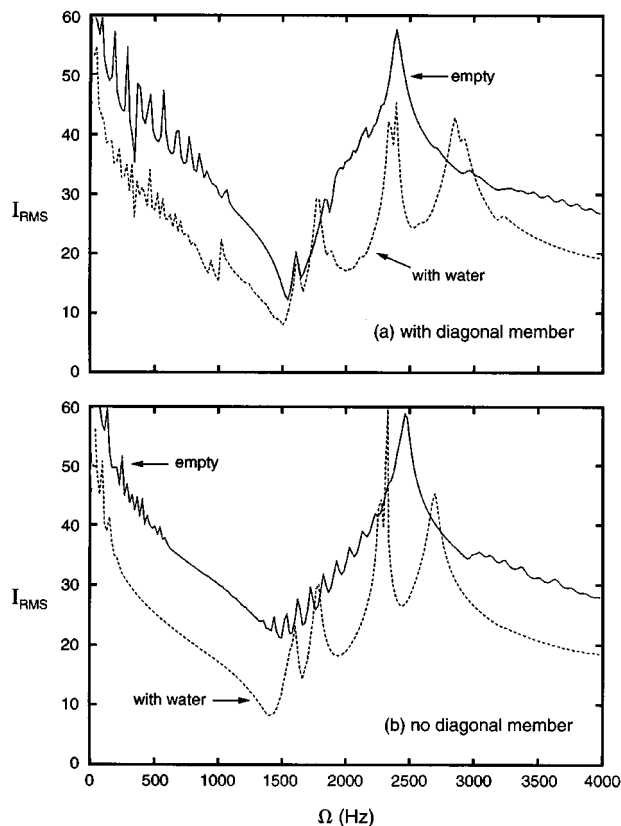


FIG. 8. Effect of water fill on frequency response, solid line: empty, dashed line: with water. $L = 2$ m, $40 \times (5 \times 4)$ cm cells, $R = \infty$.

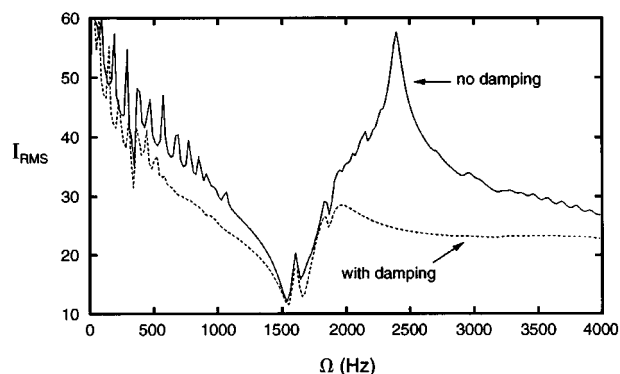


FIG. 9. Effect of damping by constrained viscoelastic layers on frequency response, solid line: no damping, dashed line: with damping. $L = 2$ m, $40 \times (5 \times 4)$ cm cells, $R = \infty$.

this zone, added mass shortens the frequency interval of propagation and attenuation zones consistent with the drop in global panel resonances and their interval from increased mass at fixed stiffness. For the panel without diagonals [Fig. 6(b)], the behavior of response curves discussed above still applies.

Figure 9 compares response of panels with and without damping treatment by constrained viscoelastic layers. The treatment consists of a damped oscillator with mass that of a rectangular polycarbonate strip and spring/dashpot that of a thin viscoelastic layer. Oscillators are bonded to corners of cells along one face of the panel (see Appendix B). In the present example, properties of the oscillator are

$$E_b = 4 \times 10^7 \text{ dyn/cm}^2, \quad E_G = 4 \times 10^8 \text{ dyn/cm}^2,$$

$$\nu_b = 0.4, \quad \tau_\epsilon = 10^{-5} \text{ s}, \quad h_b = 0.3 \text{ cm},$$

$$h_s = 1.5 \text{ cm}, \quad d_s = 2 \text{ cm}, \quad \rho_s = 1.2 \text{ g/cm}^3.$$

In the first propagation zone, damping treatment attenuates response by 8 dB. It is more effective at higher frequencies when wavelength of the panel is short. This raises relative shear motion of strip and panel causing more dissipation in the layer. The effect is greatest on the second propagation zone where peak response is attenuated by 35 dB.

III. CONCLUSION

Frequency response of a 2-D trusslike periodic panel is analyzed by treating each member of a repeated cell as a flat or circular thin strip. Transfer matrices of members in a cell are combined, yielding a cell transfer matrix. Continuity of the state vector at interfaces of cells produces the global panel transfer matrix in tridiagonal form. Noteworthy results from a parametric analysis on a typical 2-D panel are

- (1) Effect of panel curvature on response is small when the radius of curvature is larger than the wavelength of the excited mode.
- (2) Intracell flexural resonances change the nature of response from a propagation zone where the panel acts as an equivalent beam with modes having distinct wave-

lengths, to an attenuation zone where translational motion of the cell is reduced and flexural motion of each member is increased.

- (3) Flexural inertia of cell members is paramount to response, especially within attenuation zones.
- (4) Static stiffness efficiency from sag of diagonal members rises with number of cells in the panel and drops with magnitude of sag.
- (5) Frequency response is less sensitive to sag of diagonal members. Sag abates resonant peaks up to the first intra-cell flexural resonance and shifts the peak of the second propagation zone to lower frequencies.
- (6) Eliminating diagonal members in the panel has the same effect on response as sag up to the first attenuation zone.
- (7) Filling cells with water drops response by an amount proportional to added mass within the first propagation zone. Beyond this zone, added mass shortens frequency interval of propagation and attenuation zones consistent with the drop in global panel resonances and their interval from increased mass at fixed stiffness.
- (8) Damping treatment by constrained viscoelastic layers drops response in the first propagation zone. It has little effect on the first attenuation zone due to the drop in amplitude there. The effect is greatest on the second propagation zone where peak response is attenuated by 35 dB, consistent with short wavelength and higher curvature of response modes in this zone.

APPENDIX A: TRANSFER MATRIX OF MEMBER

Figure A1 shows a circular member with end points 1 and 2. Let R_m be its radius of curvature, h thickness, and $\mathbf{S} = \{f_x, f_y, m_z, u_x, u_y, \theta_z\}^T$ the state vector in member local coordinates at station s along the arc length. \mathbf{S} satisfies the equilibrium equation

$$\frac{\partial f_x}{\partial s} - \frac{f_y}{R_m} = \rho h \ddot{u}_x, \quad (\text{A1a})$$

$$\frac{\partial f_y}{\partial s} + \frac{f_x}{R_m} = \rho h \ddot{u}_y, \quad (\text{A1b})$$

$$\frac{\partial m_z}{\partial s} + f_y = \rho \frac{h^3}{12} \ddot{\theta}_z, \quad (\text{A1c})$$

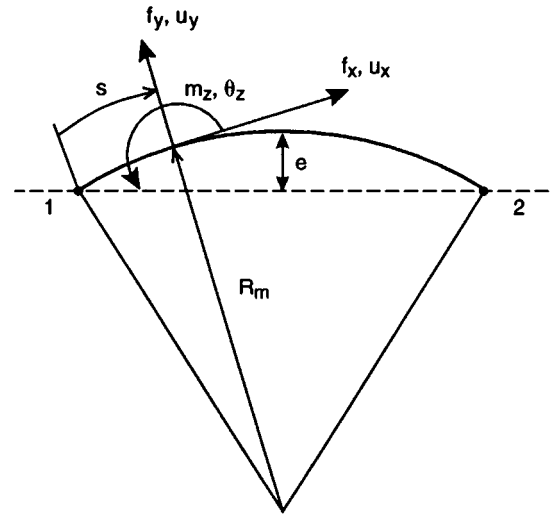


FIG. A1. Circular member.

where ρ is density and (\cdot) is derivative with respect to time. The constitutive relations are

$$\begin{aligned} f_x &= Eh \int_h \frac{\epsilon_x + y \kappa_z}{1 + y/R_m} dy, \\ \epsilon_x &= \frac{\partial u_x}{\partial s} - \frac{u_y}{R_m}, \\ \kappa_z &= \frac{\partial \theta_z}{\partial s} \Rightarrow f_x = EZ_1 \left(\frac{\partial u_x}{\partial s} - \frac{u_y}{R_m} \right) + EZ_2 \frac{\partial \theta_z}{\partial s}, \end{aligned} \quad (\text{A2})$$

$$Z^* = \frac{R}{h} \ln \left(\frac{R_m/h + 1/2}{R_m/h - 1/2} \right) - 1, \quad Z_1 = h(1 + Z^*),$$

$$Z_2 = -R_m h Z^*,$$

$$f_y = \kappa G h \left(\frac{\partial u_y}{\partial s} + \frac{u_x}{R_m} - \theta_z \right), \quad (\text{A3})$$

$$m_z = EZ_2 \left(\frac{\partial u_x}{\partial s} - \frac{u_y}{R_m} - R_m \frac{\partial \theta_z}{\partial s} \right), \quad (\text{A4})$$

where κ is shear constant. For periodic motions in time with frequency ω , (A1)→(A4) cast in matrix form are

$$\bar{\mathbf{B}} \frac{\partial \mathbf{S}}{\partial s} = \mathbf{B} \mathbf{S}, \quad (\text{A5})$$

$$\mathbf{B} = \begin{bmatrix} 0 & 1/R_m & 0 & -\rho h \omega^2 & 0 & 0 \\ -1/R_m & 0 & 0 & 0 & -\rho h \omega^2 & 0 \\ 0 & -1 & 0 & 0 & 0 & -\rho h^3 \omega^2 / 12 \\ 1/(EZ_1) & 0 & 0 & 0 & 1/R_m & 0 \\ 0 & 1/(\kappa G h) & 0 & -1/R_m & 0 & 1 \\ 0 & 0 & -1/(EZ_2 R_m) & 0 & -1/R_m^2 & 0 \end{bmatrix},$$

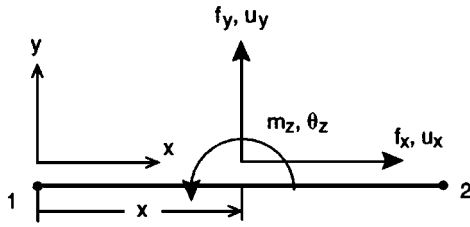


FIG. A2. Flat member.

$$\bar{\mathbf{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & +Z_2/Z_1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1/R_m & 0 & 1 \end{bmatrix}.$$

Equation (A5) admits a solution in terms of exponentials

$$\mathbf{S}(s) = \mathbf{B}_0 \mathbf{e}^{\lambda s}. \quad (\text{A6})$$

Substituting (A6) in (A5) yields

$$[\bar{\mathbf{B}}^{-1} \mathbf{B} - \lambda \mathbf{I}] \mathbf{B}_0 = 0. \quad (\text{A7})$$

A nontrivial solution of (A7) yields the dispersion relation

$$\det[\bar{\mathbf{B}}^{-1} \mathbf{B} - \lambda \mathbf{I}] = 0. \quad (\text{A8})$$

In terms of the characteristic roots λ_i , the solution of (A6) takes the form

$$\mathbf{S}(s) = \mathbf{B}_0 [\mathbf{e}^{\lambda s}] \mathbf{C}, \quad (\text{A9})$$

where \mathbf{B}_0 is the matrix of eigenvectors of (A8) and \mathbf{C} is a vector of unknown constant coefficients. Evaluating (A9) at $s=0$ and $s=l$ then eliminating \mathbf{C} yields the transfer matrix of the member

$$\mathbf{S}(l) = \mathbf{T}(l \rightarrow 0) \mathbf{S}(0) \quad (\text{A10})$$

$$\mathbf{T}(l \rightarrow 0) = \mathbf{B}_0 [\mathbf{e}^{\lambda l}] \mathbf{B}_0^{-1}.$$

The derivation above can be applied to the flat member (see Fig A2) in the limit when $R_m \rightarrow \infty$. In this case, Z_1 and Z_2 in Eq. (A2) become

$$\lim_{R \rightarrow \infty} Z_1 = h, \quad \lim_{R \rightarrow \infty} Z_2 = 0, \quad \lim_{R \rightarrow \infty} Z_2 R_m = -h^2/12.$$

APPENDIX B: IMPEDANCE AT A CELL CORNER POINT

Assume that the effect of a liquid filling the cavity of a cell is to produce translational inertia and that acoustic waves from vibratory motion of the cell walls resonate at frequencies outside the range of concern. In this way, mass of the fluid is discretized by concentrated masses acting at corner points of a cell. The stiffness matrix from liquid inertia termed \mathbf{z}_{fi} is given by

$$\mathbf{z}_{fi} = \begin{bmatrix} -m_f \frac{\omega^2}{2} & 0 & 0 \\ 0 & -m_f \frac{\omega^2}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (\text{B1})$$

where $m_f = \rho_f h_c w_c$ is mass of the liquid filling a cell, ρ_f is its density and (h_c, w_c) are height and width of a cell. Rotational inertia of the liquid driven by rotation of a cell is confined to a thin layer adjacent to the cell walls with thickness proportional to $\sqrt{\bar{\nu}_f/\omega}$ where $\bar{\nu}_f$ is kinematic viscosity of the liquid. This film is vanishingly thin for water above 10 Hz.

Another part of the impedance concerns a damping treatment by means of a constrained viscoelastic layer. This consists of a three degree of freedom oscillator with mass that of a rigid rectangular strip and spring-dashpot that of a thin viscoelastic layer as shown in Fig. B1. The damped oscillator is bonded at corners of a cell. Relative motion of the panel and strip produces irreversible transformation of mechanical energy to heat through viscoelastic straining of the layer. Let (h_s, d_s) be height and width of the strip with density ρ_s , h_b be thickness of the layer made of a viscoelastic material approximated by a linear viscoelastic solid with constitutive law

$$\sigma_b + \tau_\epsilon \dot{\sigma}_b = E_b (\epsilon_b + \tau_\sigma \dot{\epsilon}_b), \quad (\text{B2})$$

where (σ_b, ϵ_b) are stress and strain in a uniaxial strain condition; $E_b \epsilon$ is the rubbery modulus

$$E_{b\epsilon} = \frac{E_b (1 - \nu_b)}{(1 + \nu_b)(1 - 2\nu_b)}, \quad (\text{B3})$$

ν_b is Poisson ratio, E_b is Young's modulus in uniaxial stress, $(\tau_\epsilon, \tau_\sigma)$ are time constants of creep and relaxation (see Ref. 10), and (\cdot) is time derivative. For periodic motions in time with frequency ω , (B2) simplifies to

$$\sigma_b = \frac{(1 + i\omega\tau_\sigma)}{(1 + i\omega\tau_\epsilon)} E_{b\epsilon} \epsilon_b = E_c \epsilon_b, \quad (\text{B4})$$

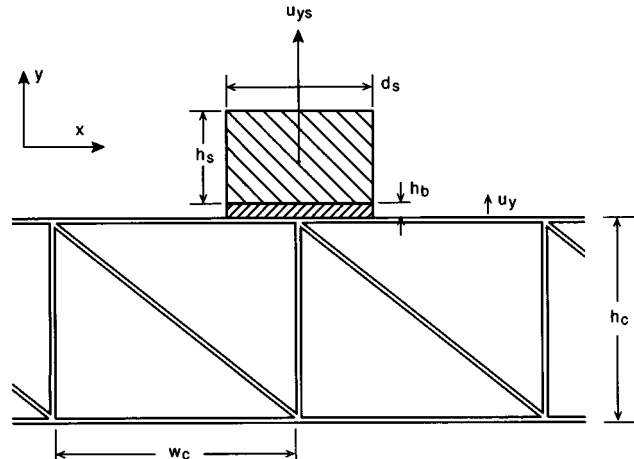


FIG. B1. Constrained viscoelastic layer.

where $i = \sqrt{-1}$ and E_c is the complex modulus. In the limit as $\omega \rightarrow \infty$

$$\lim_{\omega \rightarrow \infty} \sigma_b = E_{G\epsilon} \epsilon_b, \quad (\text{B5})$$

where $E_{G\epsilon}$ is the glassy modulus in uniaxial strain. Letting $\omega \rightarrow \infty$ in (B4) yields

$$E_{G\epsilon} = E_{b\epsilon} \frac{\tau_\sigma}{\tau_\epsilon} \Rightarrow \tau_\sigma = \frac{E_{G\epsilon}}{E_{b\epsilon}} \tau_\epsilon. \quad (\text{B6})$$

For extensional motions along Y

$$-\omega^2 m_s u_{Ys} + k_{bY}(u_{Ys} - u_Y) = 0, \quad (\text{B7})$$

$$m_s = \rho_s h_s d_s, \quad k_{bY} = E_c d_s / h_b,$$

where u_{Ys} is displacement of the strip along Y . The force exerted by the oscillator on the panel is

$$F_{sY} = -k_{bY}(u_{Ys} - u_Y) = k_{bY} \frac{\omega^2}{(\omega^2 - \omega_{sY}^2)} u_Y = z_{sY} u_Y, \quad (\text{B8})$$

$$\omega_{sY}^2 = \frac{k_{bY}}{m_s}, \quad z_{sY} = k_{bY} \frac{\omega^2}{(\omega^2 - \omega_{sY}^2)}.$$

Similarly, for shear motions along X

$$F_{sX} = -k_{bX} \frac{\omega^2}{(\omega^2 - \omega_{sX}^2)} u_X = z_{sX} u_X, \quad (\text{B9})$$

$$\omega_{sX}^2 = \frac{k_{bX}}{m_s}, \quad k_{bX} = \frac{G_c d_s}{h_b}, \quad G_c = \frac{E_c}{2(1 + \nu_b)},$$

$$z_{sX} = k_{bX} \frac{\omega^2}{(\omega^2 - \omega_{sX}^2)}.$$

Finally, for rotational motions about Z

$$M_{sZ} = k_{bZ} \frac{\omega^2}{(\omega^2 - \omega_{sZ}^2)} \theta_Z = z_{sZ} \theta_Z, \quad (\text{B10})$$

$$\omega_{sZ}^2 = \frac{k_{bZ}}{I_{ZZ}}, \quad k_{bZ} = \frac{E_c d_s^3}{12 h_b}, \quad I_{ZZ} = \frac{m_s}{12} (d_s^2 + 4 h_s^2),$$

$$z_{sZ} = k_{bZ} \frac{\omega^2}{(\omega^2 - \omega_{sZ}^2)}.$$

Combining (B8)→(B10) yields the stiffness matrix from the damping treatment

$$\mathbf{z}_{s1} = \begin{bmatrix} z_{sX} & 0 & 0 \\ 0 & z_{sY} & 0 \\ 0 & 0 & z_{sZ} \end{bmatrix}. \quad (\text{B11})$$

The total impedance from all effects at corner i is

$$\mathbf{z}_i = \mathbf{z}_{fi} + \mathbf{z}_{si}. \quad (\text{B12})$$

APPENDIX C: TRANSFER MATRIX OF A "TRUE TRUSS" MEMBER

A "true truss" member is defined as a member with infinite flexural stiffness and finite axial stiffness, hinged at

both ends. The force vector $\{f_x, f_y\}$ at an end has an axial component f_x from extensional motion and a transverse component f_y from reactions caused by rigid body translational and rotational inertias. Let subscripts 1 and 2 denote end 1 and 2 of a member. Conservation of linear and angular momentum yields equilibrium of forces and moments

$$f_{y2} + f_{y1} = -\frac{1}{2} \rho h l \omega^2 (w_2 + w_1), \quad (\text{C1})$$

$$f_{y2} - f_{y1} = -\frac{1}{6} \rho h l \omega^2 (w_2 - w_1), \quad (\text{C2})$$

where ρ is density, (h, l) are thickness and length of a flat member, and ω is radian frequency.

Adding and subtracting (C1) and (C2) yields the transfer matrix of the "true truss" member in the transverse direction

$$\begin{Bmatrix} f_y \\ w \end{Bmatrix}_2 = \begin{bmatrix} 2 & +\frac{1}{2} \rho h l \omega^2 \\ -\frac{6}{\rho h l \omega^2} & -2 \end{bmatrix} \begin{Bmatrix} f_y \\ w \end{Bmatrix}_1. \quad (\text{C3})$$

- ¹D. J. Mead, "Vibration response and wave propagation in periodic structures," ASME J. Eng. Ind., Ser. B **93**, 783–792 (1972).
- ²D. J. Mead, "A general theory of harmonic wave propagation in linear periodic systems with multiple coupling," J. Sound Vib. **27**, 235–260 (1973).
- ³D. J. Mead, "Wave propagation and natural modes in periodic systems. I. Monocoupled systems," J. Sound Vib. **40**, 1–18 (1975).
- ⁴D. J. Mead, "Wave propagation and natural modes in periodic systems: II. Multicoupled systems, with and without damping," J. Sound Vib. **40**, 19–39 (1975).
- ⁵D. J. Mead and S. Parthan, "Free wave propagation in two-dimensional periodic plates," J. Sound Vib. **64**, 325–348 (1979).
- ⁶D. J. Mead and S. Markus, "Coupled flexural-longitudinal wave motion in periodic beam," J. Sound Vib. **90**, 1–24 (1983).
- ⁷M. G. Faulkner and D. P. Hong, "Free vibrations of a mono-coupled periodic system," J. Sound Vib. **99**, 29–42 (1985).
- ⁸R. C. Engles, "Response of infinite periodic structures," J. Sound Vib. **69**, 181–197 (1980).
- ⁹A. H. vonFlotow, "Disturbance propagation in structural networks," J. Sound Vib. **106**, 433–450 (1986).
- ¹⁰Y. Yong and Y. K. Lin, "Propagation of decaying waves in periodic and piece-wise periodic structures," J. Sound Vib. **129**, 99–118 (1989).
- ¹¹J. Signorell and A. H. vonFlotow, "Wave propagation, power flow, and resonance in a truss beam," J. Sound Vib. **126**, 127–144 (1988).
- ¹²Y. Yong and Y. K. Lin, "Dynamics of complex truss-type space structures," Am. Inst. Aeron. Astron. **28**, 1250–1258 (1990).
- ¹³Y. Yong and Y. K. Lin, "Dynamic response analysis of truss-type structural networks: a wave propagation approach," J. Sound Vib. **156**, 27–45 (1992).
- ¹⁴M. El-Raheb and P. Wagner, "Damped response of shells by a constrained viscoelastic layer," ASME J. Appl. Mech. **53**, 902–908 (1986).
- ¹⁵M. El-Raheb and P. Wagner, "Damped of toroidal shells by constrained viscoelastic layer," J. Acoust. Soc. Am. **93**, 324–325 (1993).
- ¹⁶A. K. Noor and M. S. Anderson, "Continuum model for beam- and plate-like lattice structures," Am. Inst. Aeron. Astron. **16**, 1219–1228 (1978).
- ¹⁷C. T. Sun, B. J. Kim, and J. L. Bogdanoff, "On the derivation of equivalent simple models for beam- and plate-like structures in dynamic analysis," Proceedings of the 1981 AIAA Dynamics Specialists Conference (1981); pp. 523–532.

Multiple-scattering theory for mean responses in a plate with sprung masses

Richard L. Weaver

Department of Theoretical and Applied Mechanics, University of Illinois, Urbana, Illinois 61801

(Received 15 July 1996; accepted for publication 13 January 1997)

Diagrammatic multiple-scattering theory is applied to the case of an infinite homogeneous plate in flexure attached to a random distribution of sprung masses. This system is a prototypical example of a wave-bearing master structure with a locally reacting “fuzzy” substructure. Results for mean fields are obtained from the first-order smoothing approximation, the Foldy average t -matrix approximation, and Soven’s coherent-potential approximation. The study of mean-square responses is reserved to a later paper. It is found that the attenuation as calculated by Pierce *et al.* [J. Vib. Acoustics **117**, 339–348 (1995)] differs from that of the multiple-scattering theory by a fractional amount which is small if the individual sprung masses are weak. © 1997 Acoustical Society of America. [S0001-4966(97)02405-3]

PACS numbers: 43.40.Dx, 43.40.Hb [CBB]

INTRODUCTION

That complex many degree of freedom substructures can act as dynamic vibration absorbers, over a range of frequencies, to simple master structures has been attracting significant attention in recent years. In the parlance of many in the field the substructure is often termed a “fuzzy substructure.” In the limit that the number of degrees of freedom in the substructure becomes infinite, it was shown by Pierce *et al.*¹ that the chief effect of the substructure is equivalent to that of a frequency dependent viscous resistance element. The behavior of a substructure with an only finite number of degrees of freedom is, however, not as simple. Strasberg and Feit² nevertheless argued that Pierce *et al.*’s expression for the effective damping is retained for a finite degree of freedom substructure, if there is sufficient damping in the substructure to assure modal overlap. Lyon³ has analyzed similar systems and shown that the effective damping may also be derived using statistical energy analysis, thereby emphasizing that it may be understood as a consequence of flow of energy out of the master and into the substructure. Ruckman⁴ and Ruckman and Feit⁵ have provided reviews and tutorials recently in this area. In all this work there has been very little attention directed to the important case of a finite degree of freedom substructure with insufficient damping to assure modal overlap.

A simple example of such a master structure/fuzzy substructure with a single degree of freedom undamped master oscillator and a finite number of undamped degrees of freedom in the substructure was studied recently by Weaver.^{6,7} He showed analytically and numerically that Pierce *et al.*’s expression for effective damping is substantially correct even for undamped and finite degree of freedom substructures, but only until the time at which the energy achieves an equilibrium balance between master and substructure.

Pierce *et al.*¹ chiefly discussed these ideas for a continuous system consisting of a plate in bending attached to a large number of sprung masses. They argued that in the limit of a very large density of attached oscillators, the plate acts as if it were smoothly viscously damped, with a damping rate

proportional to the spectral and areal density of sprung masses and independent of the actual damping in the substructure. They also discussed a generally less interesting modification of the wave speed. There remains, however, an important question as to how this picture is modified if the density of attached oscillators is not large. The results from a single degree of freedom master^{6,7} suggest that the Pierce *et al.*¹ result is correct even for undamped substructures, but only for a finite time. After some dwell time the energy originally absorbed by the substructure should be returned to the master.

Independent work by Belyaev and Palmov^{8–10} over the past 20 years has considered wave propagation in complex continua in which there are internal degrees of freedom. This literature has chiefly emphasized the limit of a high density of internal degrees of freedom and derived the associated modifications of wave speed and dissipation. Here also, there has been little consideration of the important case in which the number density of internal degrees of freedom is finite and the damping is weak.

It is the goal of the present paper to understand the limits of the Pierce–Sparrow–Russell (PSR) theory when applied to a plate which has only a finite density of sprung masses. It is recognized that such systems are not necessarily well-described by specification of a particular response, but rather by the statistics of the responses. In the present paper we present a calculation of an ensemble average Green’s function, with particular attention directed towards the attenuation of the mean wave. Calculations of the mean-square response, and thus questions of response variance and energy flow, are reserved for a separate communication. The methods to be used are those of the well-developed theory of multiple scattering of waves in random media.

The problem is formally defined in the next section. Section II applies multiple-scattering theory to evaluate the ensemble average Green’s function. Comparisons are made in Sec. III between the PSR expression for G and the multiple-scattering expressions for $\langle G \rangle$. The predictions of the vari-

ous approaches for the attenuation of a plane wave are compared.

I. PROBLEM FORMULATION

A. Governing equations

We consider the Green's function for a flexural wave in an infinite (area $A \rightarrow \infty$) statistically homogeneous plate with a uniform areal mass density m and bending stiffness D_B . The plate is attached to an uncorrelated random distribution of a large number ($N \rightarrow \infty$) sprung masses ($g=1,2,3,\dots,N$), with finite areal number density $\rho=N/A$, having an uncorrelated distribution of spring stiffnesses k_g and masses $\mu_g = k_g/\omega_g^2$. Normalizing the Green's function to the case of a concentrated impulse of strength m , the governing equations are

$$D_B \nabla^4 G(\mathbf{x}, \mathbf{x}', t) + m \partial_t^2 G(\mathbf{x}, \mathbf{x}', t) + \sum_g k_g [G(\mathbf{x}, \mathbf{x}', t) - \psi_g(t)] \delta^2(\mathbf{x} - \mathbf{x}_g) = m \delta^2(\mathbf{x} - \mathbf{x}') \delta(t) \\ \mu_g \partial_t^2 \psi_g(t) + k_g [\psi_g(t) - G(\mathbf{x}_g, \mathbf{x}', t)] = 0, \quad (1)$$

where ψ_g represents the absolute displacement of the g th sprung mass. The solution G is the plate's displacement response to a unit concentrated impulse of magnitude m applied at point \mathbf{x}' . $\delta^2(\mathbf{x})$ is the two-dimensional Dirac delta function.

The displacements, ψ_g , of the sprung masses may be eliminated by Fourier transforming according to

$$G(\mathbf{x}, \mathbf{x}', \omega) = \int_0^{+\infty} G(\mathbf{x}, \mathbf{x}', t) \exp\{-i\omega t\} dt, \quad (2) \\ \psi_g(\omega) = \int_0^{+\infty} \psi_g(t) \exp\{-i\omega t\} dt.$$

These Fourier transforms are analytically continuable to the lower half-plane; so all poles must occur in the upper half plane. After eliminating ψ , a single partial differential equation for G results

$$D_B \nabla^4 G(\mathbf{x}, \mathbf{x}', \omega) - m \omega^2 G(\mathbf{x}, \mathbf{x}', \omega) - \left[\sum_g \frac{\omega^2 k_g \delta^2(\mathbf{x} - \mathbf{x}_g)}{\omega_g^2 - \omega^2} \right] G(\mathbf{x}, \mathbf{x}', \omega) = m \delta^2(\mathbf{x} - \mathbf{x}'). \quad (3)$$

Upon dividing through by the areal mass density the governing equation becomes

$$D \nabla^4 G(\mathbf{x}, \mathbf{x}', \omega) - \omega^2 G(\mathbf{x}, \mathbf{x}', \omega) - V(\mathbf{x}; \omega) G(\mathbf{x}, \mathbf{x}', \omega) = \delta^2(\mathbf{x} - \mathbf{x}'), \quad (4)$$

where D is bending stiffness divided by areal mass density, $D = D_B/m$. It is to this equation that the methods of multiple-scattering theory shall be applied. Due to the similar role played by V in the above equation and by potential energy in the Schrodinger equation, V may be called a "scattering potential." It differs from conventional scattering potentials in that it has a frequency dependence. Here, V is given by

$$V(\mathbf{x}; \omega) = \sum_{g=1}^N v_g(\mathbf{x}), \quad (5) \\ v_g(\mathbf{x}) = \frac{\delta^2(\mathbf{x} - \mathbf{x}_g) \omega^2 k_g}{(\omega_g^2 - \omega^2)} \equiv \delta^2(\mathbf{x} - \mathbf{x}_g) \phi(\omega_g),$$

which serves to define the quantity ϕ . The unitalicized k_g represents physical spring stiffness k_g divided by areal mass density: $k_g = k_g/m$. Similarly the unitalicized μ_g represents physical mass μ_g divided by areal plate mass density: $\mu_g = \mu_g/m$. As in the previous work,^{6,7} the springs are henceforth taken to be all equal; $k_g = k$.

B. System statistics

The sprung masses are chosen here to be distributed uniformly in space without correlations, with a number density $\rho = N/A$. Their resonant frequencies are also chosen to be uncorrelated, with a distribution

$$p(\omega_g) = \frac{\omega_0/\pi}{\omega_0^2 + \omega_g^2} \quad (-\infty < \omega_g < \infty), \quad (6)$$

where ω_0 is a characteristic frequency in the substructure. This form for p allows the needed averages to be performed in closed form. The spectral and areal mass density in the substructure is thus given by

$$\frac{dM(\omega)}{d\omega} = 2p(\omega) \rho \left[\frac{k}{\omega^2} \right] = \rho \frac{2k\omega_0/\pi\omega^2}{\omega_0^2 + \omega^2} \quad (0 < \omega < \infty). \quad (7)$$

The spectral and areal modal density in the substructure is

$$\mathcal{N}_{\text{substructure}} = 2p(\omega) \rho \quad (0 < \omega < \infty). \quad (8)$$

It may or may not be comparable to the modal density $\mathcal{N}_{\text{plate}}$ in the plate:

$$\mathcal{N}_{\text{plate}}(\omega) = \frac{1}{4\pi\sqrt{D}} = \frac{1}{4\pi} \sqrt{\frac{m}{D_B}}. \quad (9)$$

This choice for the distribution of resonant frequencies in the substructure is purposefully smooth. The very interesting case of a sharply peaked distribution is of substantial theoretical interest.¹¹ Indeed some circumstances¹² can lead to a frequency band in which modal density vanishes and wave propagation is strongly evanescent. That case is outside the scope of the present paper.

C. Dimensionless ratios

It is useful to anticipate a few relevant dimensionless ratios. Chief amongst the quantities of interest is the decay rate (inverse time) of the energy of the PSR field, given by Eq. 37 of Ref. 1, a quantity proportional to the spectral and areal density of sprung mass:

$$\sigma_{\text{PSR}} = \frac{\pi\omega^2}{2m} \frac{dM(\omega)}{d\omega} = \pi\rho k p(\omega) = \frac{\rho k \omega_0}{\omega^2 + \omega_0^2}, \quad (10) \\ \xi_{\text{PSR}} \equiv \frac{\sigma_{\text{PSR}}}{\omega}.$$

An effective dimensionless loss tangent for waves in the plate in the PSR theory is provided by dividing σ_{PSR} by frequency ω ; $\xi_{\text{PSR}} = \sigma_{\text{PSR}}/\omega$. In the applications of present interest the ratio can generally be expected to be small, corresponding to an effective loss tangent less than unity and a moderate amount of attenuation per wavelength.

We also anticipate that there is a decay rate (inverse time) of the energy in an isolated sprung mass attached to an infinite homogeneous plate. In the absence of true losses, this decay is solely due to reradiation back into the plate. It is shown later that the rate of reradiation is

$$\sigma_{\text{ism}} = k/8\sqrt{D} \equiv 1/t_{\text{dwell}}, \quad \xi_{\text{ism}} \equiv \sigma_{\text{ism}}/\omega, \quad (11)$$

which defines the dwell time. An effective loss tangent for the energy in the substructure is provided by dividing σ_{ism} by frequency; $\xi_{\text{ism}} = \sigma_{\text{ism}}/\omega$. In the usual limit¹ in which one expects a large number density of sprung masses, but a moderate value of attenuation per wavelength, ξ_{ism} would be expected to be small, corresponding to long dwell times in the substructure, and to weak springs and small individual masses.

The ratio of these two decay rates is a dimensionless quantity η ,

$$\frac{\sigma_{\text{ism}}}{\sigma_{\text{PSR}}} \equiv \eta = \frac{(\omega^2 + \omega_0^2)}{8\sqrt{D}\rho\omega_0} = \frac{1}{\rho\pi p(\omega)8\sqrt{D}}. \quad (12)$$

It is small in the usual limit invoked by Pierce *et al.*¹ that there is a large number density of small weak sprung masses. The dimensionless ratio η is the key parameter for quantifying conformity with the usual assumptions. Comparisons to be made between the predictions of the PSR theory and those of multiple-scattering theory will be correlated with this parameter, and with the loss tangents ξ_{PSR} and ξ_{ism} .

The ratio η is also the ratio of modal density in the plate to the modal density in the substructure

$$\frac{\mathcal{N}_{\text{plate}}}{\mathcal{N}_{\text{substructure}}} = \frac{1/4\pi\sqrt{D}}{2p(\omega)\rho} = \eta. \quad (13)$$

The equality of the two ratios is a consequence of reciprocity.

D. Pierce–Sparrow–Russell theory

In the limit that the number density of sprung masses, ρ , is large it is reasonable to suppose that the sum in Eq. (3) may be replaced with an integral.¹ Such a replacement is only correct for smooth functions, and this function consists of distinct resonances and is highly discontinuous. As argued by Strasberg and Feit,² and by Weaver,⁶ however, if one inserts a small amount of damping, of the order the resonance spacing, the frequency dependence in the function becomes smooth. If the number density is high as well, then the sum, it would appear, may legitimately be replaced by an integral. As emphasized by Weaver⁶ the insertion of a small amount of damping has no effect at early times, so the smoothing of the frequency dependence is legitimate if interest is confined to short times. On making this replacement Eq. (4) becomes

$$\begin{aligned} D\nabla^4 G(\mathbf{x}, \mathbf{x}', \omega) - \omega^2 G(\mathbf{x}, \mathbf{x}', \omega) \\ - \frac{\rho}{m} \left[\int_0^\infty \frac{1}{\omega_g^2 - (\omega - i\epsilon)^2} \frac{dM(\omega_g)}{d\omega_g} d\omega_g \right] G(\mathbf{x}, \mathbf{x}', \omega) \\ = \delta^2(\mathbf{x} - \mathbf{x}'). \end{aligned} \quad (14)$$

The integral over ω_g may be done in closed form

$$\begin{aligned} D\nabla^4 G(\mathbf{x}, \mathbf{x}', \omega) - \omega^2 G(\mathbf{x}, \mathbf{x}', \omega) + \rho k \left[\frac{\omega^2 + i\omega\omega_0}{\omega^2 + \omega_0^2} \right] G(\mathbf{x}, \mathbf{x}', \omega) \\ = \delta^2(\mathbf{x} - \mathbf{x}'), \end{aligned} \quad (15)$$

in which we see an effective frequency-dependent dissipation given by the imaginary part of the coefficient and a modification of the wave speed given by the real part.

The imaginary part of the coefficient may be absorbed into the ω^2 term:

$$\begin{aligned} D\nabla^4 G(\mathbf{x}, \mathbf{x}', \omega) - \left(\omega - i \frac{\rho k \omega_0}{2(\omega^2 + \omega_0^2)} \right)^2 G(\mathbf{x}, \mathbf{x}', \omega) \\ + \left[\frac{\rho k \omega^2}{\omega^2 + \omega_0^2} - \frac{\rho^2 k^2 \omega_0^2}{4(\omega^2 + \omega_0^2)^2} \right] G(\mathbf{x}, \mathbf{x}', \omega) = \delta^2(\mathbf{x} - \mathbf{x}'). \end{aligned} \quad (16)$$

The substructure has modified the plate Green's function by inducing an increased effective dynamic stiffness in the last term. It has also induced an extra damping, in which the amplitudes of a bandlimited process at frequency ω diminish with an exponential factor $\sim \exp\{-\rho k \omega_0 t / 2(\omega_0^2 + \omega^2)\}$. The PSR energy decay rate (10) is obtained from this.

The replacing of the sum with an integral is, if ρ is finite, the problematic step in this derivation. It is the intent of this paper to explore an alternative approach based on multiple-scattering theory for the mean responses and thereby to understand the limits imposed by the PSR assumption $\eta \ll 1$ and the consequences of finite η .

II. DIAGRAMMATIC MULTIPLE-SCATTERING THEORY FOR THE MEAN GREEN'S FUNCTION

A. The bare Green's function, G^0

The above equation for G must first be rewritten as an integral Lippmann–Schwinger equation in terms of the “bare” Green's function G^0 of the plate. Here, G^0 is the solution of (4) at $V=0$:

$$D\nabla^4 G^0(\mathbf{x}, \mathbf{x}', \omega) - \omega^2 G^0(\mathbf{x}, \mathbf{x}', \omega) = \delta^2(\mathbf{x} - \mathbf{x}'). \quad (17)$$

G^0 is well known in terms of Hankel and modified Hankel functions. We shall, however, be particularly interested in G^0 at $\mathbf{x}=\mathbf{x}'$. By means of spatial Fourier transforms one finds that

$$\begin{aligned}
G^0(\mathbf{0}) &= \frac{1}{4\pi^2} \int \int d^2\mathbf{q} \frac{1}{-\omega^2 + Dq^4} \\
&= \frac{1}{2\pi} \int_0^\infty q \, dq \frac{1}{-\omega^2 + Dq^4} \\
&= \frac{1}{4\pi} \int_0^\infty ds \frac{1}{-(\omega - i\epsilon)^2 + Ds^2} = \frac{-i}{8\omega\sqrt{D}}, \quad (18)
\end{aligned}$$

where an infinitesimal positive quantity ϵ has been inserted in the usual manner⁶ to resolve the pole and enforce causality and where $d^2\mathbf{q}$ is an area element $dq_1 dq_2$ in wave-vector space.

B. Diagrammatic formalism for the mean Green's function $\langle G \rangle$

The Lippmann–Schwinger equation for G in the presence of the scattering potential is obtained by convolving G^0 with both sides of (4)

$$\begin{aligned}
G(\mathbf{x}, \mathbf{x}'; \omega) &= G^0(\mathbf{x}, \mathbf{x}'; \omega) \\
&+ \int \int G^0(\mathbf{x}, \mathbf{x}'''; \omega) U(\mathbf{x}''', \mathbf{x}''; \omega) \\
&\times G(\mathbf{x}'', \mathbf{x}'; \omega) d\mathbf{x}''' d\mathbf{x}''. \quad (19)
\end{aligned}$$

In the above U is a diagonal two-component version of V : $U(\mathbf{x}, \mathbf{x}') = V(\mathbf{x}) \delta^2(\mathbf{x} - \mathbf{x}')$. Integrations over area are designated compactly by a simple $\int d\mathbf{x}$.

The Lippmann–Schwinger equation may be solved, formally, by expanding it in powers of the perturbing quantity U . In order to limit the proliferation of integration signs and variables, and to use a notation that captures only the essential topology of the equations, it is useful to recast these integral equations in terms of Feynman diagrams. A tutorial on the use of Feynman diagrams for waves in random media may be found in the review by Frisch.¹³ The diagrammatic form of Eq. (19) is

$$\text{-----} = \text{---} + \text{---} \bigcirc \text{---} \quad (20)$$

Much as in the rules of matrix multiplication, juxtapositions of symbols (Matrices) implies a contraction on adjacent indices, that is, a spatial convolution (sum) between adjacent indices \mathbf{x} . An example of such contractions is seen in the above equation (19) in which \mathbf{x}''' and \mathbf{x}'' are summed over.

On expansion in powers of U the Lippmann–Schwinger equation becomes a Neumann–Born series:

$$\text{-----} = \text{---} + \text{---} \bigcirc \text{---} + \text{---} \bigcirc \text{---} \bigcirc \text{---} + \text{---} \bigcirc \text{---} \bigcirc \text{---} \bigcirc \text{---} + \dots \quad (21)$$

or

$$\begin{aligned}
G(\mathbf{x}, \mathbf{x}') &= G^0(\mathbf{x}, \mathbf{x}') \\
&+ \int \int G^0(\mathbf{x}, \mathbf{y}) U(\mathbf{y}, \mathbf{z}) G^0(\mathbf{z}, \mathbf{x}') d\mathbf{y} d\mathbf{z} \\
&+ \int \int \int G^0(\mathbf{x}, \mathbf{y}) U(\mathbf{y}, \mathbf{z}) G^0(\mathbf{z}, \mathbf{w}) U(\mathbf{w}, \mathbf{v}) \\
&\times G^0(\mathbf{v}, \mathbf{x}') d\mathbf{y} d\mathbf{z} d\mathbf{w} d\mathbf{v} + \dots, \quad (22)
\end{aligned}$$

where the ω dependence has been suppressed. In the present case the scattering potential U consists of a number of discrete scatterers

$$\begin{aligned}
U(\mathbf{x}, \mathbf{x}'; \omega) &= \sum_{g=1}^N u_g(\mathbf{x}, \mathbf{x}'; \omega); \\
u_g(\mathbf{x}, \mathbf{x}'; \omega) &= \delta^2(\mathbf{x} - \mathbf{x}_g) \delta^2(\mathbf{x}' - \mathbf{x}_g) \phi(\omega_g)
\end{aligned} \quad (23)$$

or, in diagrammatic form,

$$\bigcirc = \sum_g \text{---} \text{---} \text{---} \quad (24)$$

The Neumann series (21) may be written in terms of the u_g :

$$\text{-----} = \text{---} + \text{---} \overset{g}{\bigcirc} \text{---} + \text{---} \overset{g}{\bigcirc} \overset{h}{\bigcirc} \text{---} + \text{---} \overset{g}{\bigcirc} \overset{h}{\bigcirc} \overset{j}{\bigcirc} \text{---} + \dots \quad (25)$$

where a summation on all scatterer indices is implied.

The chief interest is in the mean (and mean square) responses $\langle G \rangle$ and $\langle GG^* \rangle$. Thus we need to average the Neumann series over the statistics of the scatterers. To do so it is necessary to understand the correlations, if any, between the scatterers (none are assumed in the present application) and we need to consider the possibility that, in the above sums over g and h etc., that $g=h$ in some of the terms. Thus we first festoon the above diagrams with “links” indicating identity of otherwise apparently distinct scatterers:

$$\text{-----} = \text{---} + \text{---} \overset{g}{\text{---}} \overset{g}{\bigcirc} \text{---} + \text{---} \overset{g}{\text{---}} \overset{g}{\bigcirc} \overset{h}{\text{---}} \overset{h}{\bigcirc} \text{---} + \text{---} \overset{g}{\text{---}} \overset{g}{\bigcirc} \overset{g}{\text{---}} \overset{g}{\bigcirc} \text{---} + \text{---} \overset{g}{\text{---}} \overset{g}{\bigcirc} \overset{g}{\text{---}} \overset{h}{\text{---}} \overset{h}{\bigcirc} \text{---} + \text{---} \overset{g}{\text{---}} \overset{g}{\bigcirc} \overset{h}{\text{---}} \overset{h}{\bigcirc} \text{---} + \text{---} \overset{g}{\text{---}} \overset{g}{\bigcirc} \overset{h}{\text{---}} \overset{j}{\text{---}} \overset{j}{\bigcirc} \text{---} + \dots \quad (26)$$

The summations are now confined to the cases $g \neq h$, etc. At this point the scatterer labels g, h, i, \dots are no longer needed and will be dropped.

The above expression may be understood as a series of different scattering paths. For example, the fifth term on the right side may be understood as a process of propagation from the source to the g th scatterer, then to the h th, then back to the g th, and finally to the receiver. The actual diagram corresponds, because of the sum over h and g , to all possible paths that scatter off one scatterer, then off another, and then back to the first before being received.

Upon averaging, the diagrams look exactly the same, except that now one must understand the symbols \bigcirc as having been averaged. We use the symbol --- for $\langle G \rangle$ and write

$$\text{---} = \text{---} + \text{---} \text{---} \text{---} + \text{---} \text{---} \text{---} \text{---} + \text{---} \text{---} \text{---} \text{---} \text{---} + \dots \quad (27)$$

where \blacksquare is now understood as $\langle \Sigma_g u_g \rangle$. If there were correlations amongst the scatterers, then upon averaging the diagrams would gain another kind of link also¹³ and the series would be a bit more complicated.

For those who are unfamiliar with Feynman diagrams it may be helpful to write out one of the above terms. The fifth diagram on the right, for example, corresponds to the integral expression

$$\sum_{g \neq h} \int \int \int \int \int \int \int G^0(\mathbf{x}, \mathbf{y}) \mathbf{v}_g(\mathbf{y}) G^0(\mathbf{y}, \mathbf{w}) \mathbf{v}_h(\mathbf{w}) G^0(\mathbf{w}, \mathbf{z}) \mathbf{v}_g(\mathbf{z}) G^0(\mathbf{z}, \mathbf{x}') d\mathbf{y} d\mathbf{z} d\mathbf{w} \frac{d\mathbf{x}_g d\mathbf{x}_h}{A^2} p(\omega_g) p(\omega_h) d\omega_g d\omega_h$$

$$= \rho^2 \int \int \int \int G^0(\mathbf{x}, \mathbf{x}_g) \phi(\omega_g) G^0(\mathbf{x}_g, \mathbf{x}_h) \phi(\omega_h) G^0(\mathbf{x}_h, \mathbf{x}_g) \phi(\omega_g) G^0(\mathbf{x}_g, \mathbf{x}') d\mathbf{x}_g d\mathbf{x}_h p(\omega_g) p(\omega_h) d\omega_g d\omega_h, \quad (28)$$

where $N(N-1)/A^2 = \rho^2$ for large N . The efficiency of the diagram notation is apparent.

The infinite series (27) above can be formally summed by defining a “self-energy,” or “mass,” operator, m , as the sum of all uncuttable (see Frisch¹³ for a more extended discussion of uncuttability) diagrams, that is, diagrams which cannot be factored into (“matrix,” or spatial convolution) products of other diagrams.

$$\bullet = \cdot + \text{---}\square + \text{---}\square\text{---}\square + \text{---}\square \text{---}\square + \text{---}\square \text{---}\square + \dots \quad (29)$$

In terms of m the infinite series (27) is summed, exactly, into the Dyson equation:

$$\text{---}\bullet = \text{---} + \text{---}\bullet \quad (30)$$

or,

$$\langle G(\mathbf{x}, \mathbf{x}') \rangle = G^0(\mathbf{x}, \mathbf{x}') + \int \int d\mathbf{x}'' d\mathbf{x}''' G^0(\mathbf{x}, \mathbf{x}'') \times m(\mathbf{x}'', \mathbf{x}''') \langle G(\mathbf{x}''', \mathbf{x}') \rangle,$$

which may be proven by expanding it in powers of m and comparing with (27).

In a statistically homogeneous medium the self-energy operator is a function of only the difference of its arguments, and the Dyson equation is solvable by spatial Fourier transforms. In the approximations considered here m will take the especially simple form

$$m(\mathbf{x}, \mathbf{x}') = \mathbf{m}(\omega) \delta^2(\mathbf{x} - \mathbf{x}'), \quad (31)$$

where $\mathbf{m}(\omega)$ is independent of position. In this case the spatial Fourier transform procedure gives a solution $\langle G(\mathbf{x}, \mathbf{x}') \rangle$ to the Dyson equation of the form

$$G(\mathbf{x}, \mathbf{x}'; \omega) = \frac{1}{4\pi^2} \int \int d^2\mathbf{q} \frac{\exp\{-i\mathbf{q} \cdot (\mathbf{x} - \mathbf{x}')\}}{D\mathbf{q}^4 - \omega^2 - \mathbf{m}(\omega)}. \quad (32)$$

C. First-order smoothing approximation

To leading order in the scattering strengths m may be approximated by its first one or two terms.

$$\bullet \approx \cdot + \text{---}\square \quad (33)$$

$$m(\mathbf{x}, \mathbf{x}') \approx \rho \int \phi(\varpi) p(\varpi) d\varpi \delta^2(\mathbf{x} - \mathbf{x}') + \rho \int \phi^2(\varpi) p(\varpi) d\varpi G^0(\mathbf{0}) \delta^2(\mathbf{x} - \mathbf{x}').$$

A symbol ϖ has been introduced to play the role of a generic substructural resonant frequency ω_g . This approximation for m is equivalent to the first-order smoothing approximation, or FOSA,^{13,14} and is presumably valid to first order in the scatterer number density ρ and to second order in the scatterer strengths. The FOSA is more commonly presented within a context in which the scatterers have zero mean, so that the first of the above terms vanishes. Belyaev¹⁰ recently applied the FOSA, and the Dyson equation, to estimate mean responses in a context of one-dimensional structural vibrations.

D. Foldy approximation

A truncated expansion like (33) in powers of ϕ is a problematic one; ϕ is resonant, and infinitely so, at precisely the places where it is most important. One may, though, expand to all orders by considering the scattering operator, t , of a single scatterer. In terms of t one may then invoke another approximation, termed the average t -matrix approximation or Foldy approximation.¹⁵ In this approximation one takes advantage of the exact solvability of the single scatterer case

$$t_g(\mathbf{x}, \mathbf{x}') = u_g(\mathbf{x}, \mathbf{x}') + \int \int u_g(\mathbf{x}, \mathbf{x}'') \times G^0(\mathbf{x}'', \mathbf{x}''') t_g(\mathbf{x}''', \mathbf{x}') d\mathbf{x}'' d\mathbf{x}'''. \quad (34)$$

If the operator t is represented by the symbol \otimes , then this equation is represented by the diagram equation

$$\otimes = \cdot + \text{---}\square + \text{---}\square\text{---}\square + \text{---}\square \text{---}\square + \text{---}\square \text{---}\square + \dots \quad (35)$$

It has solution

$$t_g(\mathbf{x}, \mathbf{x}') = [\phi^{-1}(\omega_g) - G^0(\mathbf{0})]^{-1} \delta^2(\mathbf{x} - \mathbf{x}_g) \delta^2(\mathbf{x}' - \mathbf{x}_g) \equiv \tau(\omega_g) \delta^2(\mathbf{x} - \mathbf{x}_g) \delta^2(\mathbf{x}' - \mathbf{x}_g), \quad (36)$$

where τ has been defined by

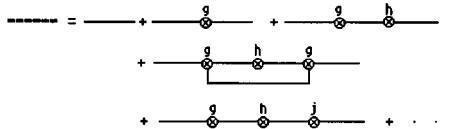
$$\tau(\omega_g) \equiv [\phi^{-1}(\omega_g) - G^0(\mathbf{0})]^{-1} = \frac{\omega^2 k}{\omega_g^2 - \omega^2 + i\omega k/8\sqrt{D}}. \quad (37)$$

We take parenthetical note of an optical theorem:

$$|\tau(\omega_g)|^2 = -\text{Im}\{\tau(\omega_g)\} 8\omega\sqrt{D}, \quad (38)$$

which may be interpreted as a relationship between total cross section and forward scattering amplitude, and as a consequence of energy conservation during scattering. The relationship would fail if there were dissipation in the oscillator.

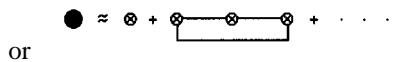
The Neumann series for G is then written in terms of t in the form:



$$G = \text{---} \circ \text{---} + \text{---} \circ \text{---} \circ \text{---} + \text{---} \circ \text{---} \circ \text{---} \circ \text{---} + \dots \quad (39)$$

In this case links cannot connect adjacent scatterers because a factor $t_g G^0 t_g$ would erroneously double count terms in the infinite sum (35). That (39) is equal to (27), except with a re-ordering of the terms, may be seen by substituting (35) into it.

Upon averaging the above series, and again defining the self-energy operator as the sum of all uncuttable diagrams:



$$\bullet \approx \circ + \text{---} \circ \text{---} \circ \text{---} + \dots \quad (40)$$

or

$$m(\mathbf{x}, \mathbf{x}') \approx \rho \int \tau(\varpi) p(\varpi) d\varpi \delta^2(\mathbf{x} - \mathbf{x}') + \rho^2 \int \int \tau^2(\varpi) \tau(\varpi') p(\varpi) p(\varpi') \times d\varpi d\varpi' G^0(\mathbf{x}, \mathbf{y}) G^0(\mathbf{y}, \mathbf{x}) d\mathbf{y} \delta^2(\mathbf{x} - \mathbf{x}') + \dots$$

one again deduces a Dyson equation (30).

A truncation of the series (40) for m at the first term may be termed the average t -matrix approximation. It is equivalent to that of Foldy.¹⁵ It may be noted that such an approximation includes both terms in the FOSA, and higher-order ones as well.

A correction to Foldy's approximation is provided by including the next term in (40). This correction is of order ρ^2 but contains only those terms of order ρ^2 which are of order τ^3 or less. Romack and Weaver¹⁶ termed this approximation the "m3 approximation." Waterman and Truell¹⁷ estimated the correction in order to assess the validity of the Foldy approximation. As t has the delta-function character indicated in (36), it is possible to evaluate the term exactly by performing the integral over \mathbf{y} . One obtains

$$m(\mathbf{x}, \mathbf{x}') \approx \left[\rho \int \tau(\varpi) p(\varpi) d\varpi \delta^2(\mathbf{x} - \mathbf{x}') \right] \left[1 + \rho \int \tau^2(\varpi) \times p(\varpi) d\varpi \int G^0(\mathbf{x}, \mathbf{y}) G^0(\mathbf{y}, \mathbf{x}) d\mathbf{y} + \dots \right]$$

$$= \left[\rho \int \tau(\varpi) p(\varpi) d\varpi \delta^2(\mathbf{x} - \mathbf{x}') \right] \left[1 + i\rho \int \tau^2(\varpi) \times p(\varpi) d\varpi / (16\omega^3 \sqrt{D}) + \dots \right]. \quad (41)$$

Thus the Waterman and Truell¹⁷ criterion for the validity of the Foldy approximation is

$$\left| \rho \int \tau^2(\varpi) p(\varpi) d\varpi / (16\omega^3 \sqrt{D}) \right| \ll 1. \quad (42)$$

E. Coherent potential approximation

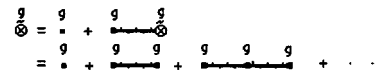
The coherent potential approximation,¹⁸ or CPA, is a self-consistent method that requires consideration of a medium, with a propagator \tilde{G} , that has a spatially homogeneous complex frequency dependent "coherent potential" χ .

$$D\nabla^4 \tilde{G}(\mathbf{x}, \mathbf{x}') - \omega^2 \tilde{G}(\mathbf{x}, \mathbf{x}') + \chi \tilde{G}(\mathbf{x}, \mathbf{x}') = \delta^2(\mathbf{x} - \mathbf{x}'). \quad (43)$$

$\tilde{G}(\mathbf{0})$ may be evaluated:

$$\begin{aligned} \tilde{G}(\mathbf{0}) &= \frac{1}{4\pi^2} \int \int d^2\mathbf{q} \frac{1}{\chi - \omega^2 + D\mathbf{q}^4} \\ &= \frac{1}{2\pi} \int_0^\infty q dq \frac{1}{\chi - \omega^2 + Dq^4} \\ &= \frac{1}{4\pi} \int_0^\infty ds \frac{1}{\chi - (\omega - i\epsilon)^2 + Ds^2} = \frac{-i}{8\sqrt{D}\sqrt{\omega^2 - \chi}}. \end{aligned} \quad (44)$$

The symbol for the propagator in the medium with a coherent potential is chosen to be $\tilde{G} = \text{---} \circ \text{---}$. The scattering operator \tilde{t} off a single scatterer in this medium is given by



$$\tilde{t} = \text{---} \circ \text{---} + \text{---} \circ \text{---} \circ \text{---} + \dots \quad (45)$$

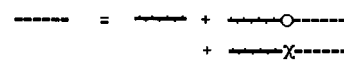
which has solution

$$\tilde{t} = \tilde{t}_g(\mathbf{x}, \mathbf{x}') = [\phi^{-1}(\omega_g) - \tilde{G}(\mathbf{0})]^{-1} \delta^2(\mathbf{x} - \mathbf{x}_g) \delta^2(\mathbf{x}' - \mathbf{x}_g). \quad (46)$$

On averaging one obtains

$$\begin{aligned} \left\langle \sum_g \tilde{t}_g(\mathbf{x}, \mathbf{x}') \right\rangle &= \sum_g \int \int \frac{d\mathbf{x}_g}{A} p(\varpi) d\varpi \tilde{t}_g(\mathbf{x}, \mathbf{x}') \\ &= \rho \int_{-\infty}^\infty p(\varpi) d\varpi [\phi^{-1}(\varpi) - \tilde{G}(\mathbf{0})]^{-1} \delta^2(\mathbf{x} - \mathbf{x}'). \end{aligned} \quad (47)$$

The Lippmann-Schwinger equation now has two sources of scattering, U , and a term χ to correct for the introduction of the coherent potential. G is given exactly by



$$G = \text{---} \circ \text{---} + \text{---} \circ \text{---} \circ \text{---} + \dots \quad (48)$$

The mass operator then includes both scatterings from the t 's and from χ .

$$\bullet = \tilde{\otimes} + \chi + \begin{array}{c} \tilde{\otimes} \quad \tilde{\otimes} \\ \hline \end{array} + \begin{array}{c} \tilde{\otimes} \quad \chi \quad \tilde{\otimes} \\ \hline \end{array} + \dots \quad (49)$$

where again the same symbol \otimes is used for both \tilde{t} and $\langle \tilde{t} \rangle$. The Dyson equation now reads:

$$\text{---} = \text{---} + \text{---} \bullet \text{---} \quad (50)$$

If one chooses χ , which so far has been a free parameter, such that

$$\tilde{\otimes} = -\chi \delta^2(\mathbf{x} - \mathbf{x}') \quad (51)$$

then the first four terms in m , and many others as well, cancel, leaving m of order $\rho^2 \tau^4 \sim 0$. It is therefore argued that the CPA should be more accurate than the Foldy approximation. For the current application the CPA approximation reduces to

$$\rho \int d\varpi p(\varpi) [\phi^{-1}(\varpi) - \tilde{G}(\mathbf{0})]^{-1} = -\chi, \quad (52)$$

where \tilde{G} and χ are also related by (44).

III. APPLICATIONS

Each of these approximations, the FOSA, the Foldy, and the CPA, may now be applied to the case of interest.

A. First-order smoothing approximation

The first term of the FOSA is

$$\begin{aligned} m_{\text{FOSA-1}} &= \delta^2(\mathbf{x} - \mathbf{x}') \rho \int_{-\infty}^{\infty} d\varpi \frac{\omega_0/\pi}{\omega_0^2 + \varpi^2} \frac{\omega^2 k}{\varpi^2 - (\omega - i\epsilon)^2} \\ &= -\frac{\delta^2(\mathbf{x} - \mathbf{x}') \rho k}{\omega_0^2 + \omega^2} [\omega^2 + i\omega\omega_0]. \end{aligned} \quad (53)$$

Substituting this in Eq. (32), one finds that

$$\begin{aligned} \langle G^{\text{FOSA-1}}(\mathbf{x}, \mathbf{x}') \rangle &= \frac{1}{4\pi^2} \int \int d^2\mathbf{q} \\ &\times \frac{\exp\{-i\mathbf{q} \cdot (\mathbf{x} - \mathbf{x}')\}}{Dq^4 - \omega^2 + [\rho k(\omega^2 + i\omega\omega_0)]/[\omega_0^2 + \omega^2]}, \end{aligned} \quad (54)$$

which has poles at effective flexural wave numbers given by

$$q^*(\omega) = \sqrt[4]{\omega^2 - \frac{\rho k}{\omega_0^2 + \omega^2 - i\omega k/8\sqrt{D}} \left(\omega^2 + \frac{i\omega\omega_0}{\sqrt{1 - ik/8\omega\sqrt{D}}} \right)} / \sqrt[4]{D}. \quad (58)$$

The decay of the energy of the mean field is given, in this approximation, by $\sigma_{\text{Foldy}} = -\text{Im } \mathbf{m}_{\text{Foldy}}/\omega$. It may be shown that the Foldy decay rate is given by

$$\sigma_{\text{Foldy}} = \sigma_{\text{PSR}} \left[1 + \xi_{\text{ism}} \frac{\omega^3}{\omega_0(\omega^2 + \omega_0^2)} \right]. \quad (59)$$

$$q^*(\omega) = \sqrt[4]{\omega^2 - \rho k \frac{\omega^2}{\omega_0^2 + \omega^2} - i\rho k \frac{\omega\omega_0}{\omega_0^2 + \omega^2}} / \sqrt[4]{D}. \quad (55)$$

This is the wave number of Pierce *et al.*¹ We note that the attached sprung masses have induced a change in the wave speed, and induced an apparent attenuation. We also note that the energy of the mean field decays, in this approximation, at a rate $\sigma_{\text{FOSA-1}} = \sigma_{\text{PSR}} = -\text{Im } \mathbf{m}/\omega = \rho\omega_0 k/(\omega_0^2 + \omega^2)$.

The second term of the FOSA is

$$\begin{aligned} m_{\text{FOSA-2}} &= \delta^2(\mathbf{x} - \mathbf{x}') G^0(\mathbf{0}) \rho \int_{-\infty}^{\infty} d\varpi \frac{\omega_0/\pi}{\omega_0^2 + \varpi^2} \\ &\times \left[\frac{\omega^2 k}{\varpi^2 - (\omega - i\epsilon)^2} \right]^2 \\ &= \frac{\delta^2(\mathbf{x} - \mathbf{x}') k^2 G^0(\mathbf{0}) \rho \omega^2}{(\omega_0^2 + \omega^2)^2} [\omega^2 + i\omega\omega_0], \end{aligned} \quad (56)$$

which differs from (53) by a factor of order ξ_{ism} .

B. Foldy approximation

The Foldy, approximation, which incorporates FOSA-1 and FOSA-2 as well as higher order terms in ϕ , is

$$\begin{aligned} m_{\text{Foldy}} &= \rho \delta^2(\mathbf{x} - \mathbf{x}') \int_{-\infty}^{\infty} d\varpi \frac{\omega_0/\pi}{\omega_0^2 + \varpi^2} \tau(\varpi) \\ &= \mathbf{m}_{\text{Foldy}} \delta^2(\mathbf{x} - \mathbf{x}') \\ &= \rho \delta^2(\mathbf{x} - \mathbf{x}') \int_{-\infty}^{\infty} d\varpi \frac{\omega_0/\pi}{\omega_0^2 + \varpi^2} \frac{\omega^2 k}{\varpi^2 - \omega^2 - \omega^2 k G^0(\mathbf{0})} \\ &= -\frac{\rho \delta^2(\mathbf{x} - \mathbf{x}') k}{\omega_0^2 + \omega^2 + \omega^2 k G^0(\mathbf{0})} \left[\omega^2 + \frac{i\omega\omega_0}{\sqrt{1 + kG^0(\mathbf{0})}} \right] \\ &= -\frac{\rho \delta^2(\mathbf{x} - \mathbf{x}') k}{\omega_0^2 + \omega^2 - i\omega k/8\sqrt{D}} \left[\omega^2 + \frac{i\omega\omega_0}{\sqrt{1 - ik/8\omega\sqrt{D}}} \right]. \end{aligned} \quad (57)$$

This corresponds to a wave number $q^*(\omega)$ given by

The PSR and Foldy decay rates differ by a fractional amount of order ξ_{ism} . A criterion for the validity of the PSR result is therefore $\xi_{ism} \ll 1$, that is, that the individual springs must be sufficiently weak. One notes that this criterion (but there may be others also) puts no constraints on their number density, except inasmuch as $\xi_{ism} \ll 1$ implies that ρ must be large if one is nevertheless to have non-negligible effects proportional to ρk .

The Waterman and Truell criterion (42) for the validity of the Foldy approximation can also be examined

$$\left| \frac{\rho \omega^4 k^2}{(\omega^2 + \omega_0^2 - i \omega k / 8 \sqrt{D})^2} \left[1 - \frac{i}{2} \frac{\omega_0}{\sqrt{\omega^2 - i \omega k / 8 \sqrt{D}}} \right] \right| \ll 16 \omega^3 \sqrt{D}. \quad (60)$$

For ω of the order of ω_0 this is essentially

$$\frac{\sigma_{PSR}}{\omega} \frac{\sigma_{ism}}{\omega} = \xi_{PSR} \xi_{ism} \ll 1, \quad (61)$$

which is well-satisfied in the envisioned applications.

C. Coherent potential approximation

The CPA differs slightly from the Foldy approximation. It requires a value of χ such that

$$\begin{aligned} -\chi &= \tilde{\tau} = \rho \int d\varpi \frac{\omega_0 / \pi}{\omega_0^2 + \varpi^2} \frac{\omega^2 k}{\varpi^2 - \omega^2 - \omega^2 k \tilde{G}(\mathbf{0})} \\ &= - \frac{\rho k}{\omega_0^2 + \omega^2 + \omega^2 k \tilde{G}(\mathbf{0})} \left[\omega^2 + \frac{i \omega \omega_0}{\sqrt{1 + k \tilde{G}(\mathbf{0})}} \right] \\ &= - \frac{\rho k}{\omega_0^2 + \omega^2 - i \omega^2 k / 8 \sqrt{D} \sqrt{\omega^2 - \chi}} \\ &\quad \times \left[\omega^2 + \frac{i \omega \omega_0}{\sqrt{(1 - i k / 8 \sqrt{D} \sqrt{\omega^2 - \chi})}} \right]. \end{aligned} \quad (62)$$

The corresponding wave number q^* is

$$q^*(\omega) = \sqrt[4]{\omega^2 - \chi} / \sqrt{D}. \quad (63)$$

The imaginary part of q^* is the attenuation of the mean field.

D. Comparisons of the predictions

The attenuation, as predicted by the different theories, is evaluated and plotted in Fig. 1. Plots versus frequency are not shown, as they would largely reflect the arbitrary choice (6) for the spectral distribution of substructural resonant frequencies. Thus the results are plotted versus ρk and k for fixed ω . The values chosen were $\omega = \omega_0 = D = 1$. We note that for all theories, and all values of k examined, the attenuation is given by a number which is very roughly $\rho k / 7$. The greatest difference between the PSR attenuations and the presumably more accurate Foldy attenuation is found at large values of ξ_{ism} and ξ_{PSR} where the PSR attenuation is nearly twice the Foldy attenuation. The PSR expression is usually an overestimate. The Foldy and CPA theories are not significantly different in this parameter range.

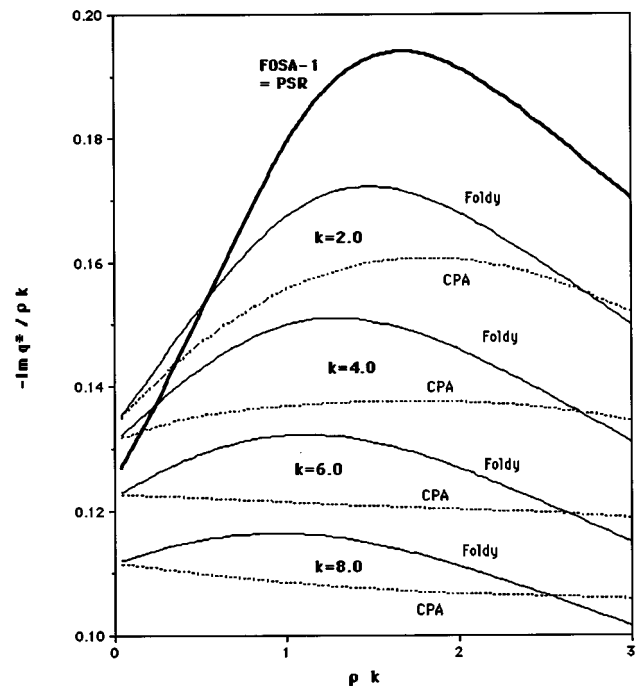


FIG. 1. The attenuation is plotted versus ρk for the three different theories for the case $\omega_0 = D = 1$, and for four different values of the spring stiffness k , corresponding to $\xi_{ism} = 0.25, 0.50, 0.75$, and 1.0 . The PSR result is plotted in bold; the Foldy results with a solid line; the CPA results with a dotted line. As k vanishes all theories approach the PSR.

Introduction of true damping mechanisms will do little to change this picture. If the loss tangents (due to true dissipation) in the plate and in the substructure are equal, the effect of such introduction is a trivial multiplication of all responses by the same exponentially decreasing factor $\exp(-\xi_{true} \omega t)$. If the losses in the substructure exceed those in the plate, then the effect of the additional damping can be incorporated most easily by letting k [in, e.g., Eq. (58)] get a slight imaginary part: $k \rightarrow k(1 - i\xi)$.

IV. CONCLUSIONS

Diagrammatic multiple-scattering theory has been applied to the case of a random distribution of sprung masses attached to an infinite plate. The resulting expressions for the ensemble averaged Green's functions, under various approximations to the multiple-scattering equations, were compared to the Green's function of the Pierce-Sparrow-Russell (PSR) effective plate. It was found that the PSR result corresponds to the zeroth-order smoothing approximation, the simplest of the multiple-scattering approximations.

It was also found that other approximations, the first-order smoothing approximation, the Foldy approximation, and the Coherent potential approximation, give results (for the speed and the attenuation of the mean plane wave) that differ from those of PSR. The differences are found to be slight if the stiffnesses of the individual sprung masses are weak compared to the stiffness of the plate; $\xi_{ism} \ll 1$. If this quantity is not small, then the differences remain slight if the areal density of the sprung masses is small; the differences are significant only if the loss tangent of the PSR plane wave

is comparable to unity. That is a parameter regime outside of the chief interest. Over the entire parameter range studied, the PSR result for attenuation was never found to exceed other predictions by more than a factor of two.

It is recognized that the mean Green's function does not necessarily fully represent responses that would be obtained in specific samples from the random ensemble. Thus it is appropriate to ask other questions as well. The mean-square Green's function includes contributions from the coherent and the incoherent parts of the field. It would allow determination of the variance around the mean, and the flow of energy in the structure. Discussion of the mean-square Green's function is reserved for a later paper.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research through Grant No. N00014-94-0855.

- ¹A. D. Pierce, V. W. Sparrow, and D. A. Russell, "Fundamental structural-acoustic idealizations for structures with fuzzy internals," *J. Vib. Acoust.* **117**, 339–348 (1995).
- ²M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," *J. Acoust. Soc. Am.* **99**, 335–344 (1996).
- ³R. H. Lyon, "Statistical energy analysis and structural fuzzy," *J. Acoust. Soc. Am.* **97**, 2878–2881 (1995).
- ⁴C. Ruckman, "A review of publications related to fuzzy structures analysis" (private communication).
- ⁵C. E. Ruckman and D. Felt, "A Tutorial on Soize's method for stochastic modeling in structural acoustics (Fuzzy Structures Analysis)," in *Proceedings of the ASME 15th Biennial Conference on Mechanical Vibration*

and Noise (American Society of Mechanical Engineers, Boston, MA, 1995), pp. 241–246.

- ⁶R. Weaver, "The effect of an undamped finite degree of freedom "fuzzy" substructure: numerical solutions and theoretical discussion," *J. Acoust. Soc. Am.* **100**, 3159–3164 (1996).
- ⁷R. Weaver, "Mean and Mean Square Responses of a Prototypical Master/Fuzzy Structure," *J. Acoust. Soc. Am.* (in press).
- ⁸A. K. Belyaev and V. A. Palmov, "Integral theories of random vibration of complex structures," in *Random Vibration—Status and Recent Developments*, edited by R. H. Lyon and I. Elishakoff (Elsevier, Amsterdam, 1986), pp. 19–38.
- ⁹A. K. Belyaev, "Vibrational state of complex mechanical structures under broad-band excitation," *Int. J. Solids Struct.* **27**, 811–823 (1991).
- ¹⁰A. K. Belyaev, "High-frequency vibration of extended complex structures," *Prob. Eng. Mech.* **8**, 15–24 (1993).
- ¹¹C. Feuillade, "The attenuation and dispersion of sound in water containing multiply interacting air bubbles," *J. Acoust. Soc. Am.* **99**, 3412–3430 (1996).
- ¹²M. Junger and J. C. Cole, "Bubble swarm acoustics: insertion loss of a layer on a plate," *J. Acoust. Soc. Am.* **68**, 241 (1980).
- ¹³U. Frisch, "Wave Propagation in Random Media," in *Probabilistic Methods in Applied Mathematics*, edited by T. Bharucha-Reid (Academic New York, 1968), Vol. 1, pp. 75–197.
- ¹⁴J. J. McCoy, "On the calculation of bulk properties of heterogeneous materials," *Q. J. Appl. Math.*, 137–149 (1979).
- ¹⁵L. L. Foldy, "The multiple scattering of waves, I. General theory of isotropic scattering by randomly distributed scatterers," *Phys. Rev.* **67**, 107–119 (1945).
- ¹⁶G. Romack and R. Weaver, "Monte Carlo study of multiple scattering of waves in two-dimensional random media," *J. Acoust. Soc. Am.* **94**, 506–513 (1993).
- ¹⁷P. Waterman and R. Truell, "Multiple scattering of waves," *J. Math. Phys. (N.Y.)* **2**, 512–537 (1961).
- ¹⁸P. Soven, "Coherent-potential model of substitutional disordered alloys," *Phys. Rev.* **85**, 621–629 (1967).

Estimating acoustic radiation from a Bernoulli–Euler beam using shaped polyvinylidene fluoride film

Brian L. Scott

Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802

Scott D. Sommerfeldt

Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602

(Received 3 July 1996; accepted for publication 8 February 1997)

This paper examines numerically the use of arrayed shaped polyvinylidene fluoride film sensors to obtain an estimate of the far-field radiated power from a clamped-clamped Bernoulli–Euler beam excited by a sinusoidal point force. Distributed sensors are designed that are sensitive to the radiated power. Fourier analysis of point and distributed sensor outputs is used to estimate the beam's radiated power. The numerical results of the estimated power obtained from the point and distributed sensors are compared to those predicted by the theoretical model and the performance of the two sensor types is assessed. The results show that the shaped sensors provide a reasonable estimate of the strain transforms over the region of interest, when compared to the exact theoretical predictions, for the first five beam resonance frequencies and for frequencies located approximately midway between resonance frequencies. The distributed sensors also give consistently better predictions of the strain transforms for all frequencies considered when compared to the point sensor results. The power predictions using shaped sensors range from -1 to 3 dB relative to the exact theoretical results for most of the frequencies analyzed. This contrasts to the 36 – 68 dB range of relative power estimates using point sensors. © 1997 Acoustical Society of America. [S0001-4966(97)02606-4]

PACS numbers: 43.40.Rj, 43.38.Fx [CBB]

INTRODUCTION

A number of methods exist to obtain the acoustic radiation from a vibrating structure. Examples include an analytical model of the structure to predict radiated power, microphones to perform far-field or intensity measurements, or accelerometers to obtain the structure's vibrational response. A known model assumption has the benefit of not requiring additional financial, weight, or space costs, but normally cannot model the structure completely and lacks the ability to adapt to changes in the system. Using transducers to collect real-time data allows the prediction of radiated power even when system parameters change. If transducers are used, the physical quantity to be measured and the location of the sensors are important considerations.

The radiation from a structure can be determined by placing transducers away from the structure, as is the case when using microphones to record the pressure levels in the surrounding fluid. Unattached sensors have the benefit of having little or no effect on the behavior of the system under consideration and measure only what is radiated when implemented properly. However, locating transducers off the structure is often not suitable due to environmental constraints. Mounting sensors directly on the structure, for instance by using accelerometers, can solve this problem. In addition, this type of sensor is not influenced by secondary pathways and might be better for implementing active control of the structure. Since the desire is to determine the radiated power, transducers that are sensitive to radiation, not simply vibration, are required. One possibility for

achieving this is to use distributed sensors mounted to the structure.

Using distributed sensors allows the fabrication of sensors that are sensitive to specific vibrational characteristics of the structure. Work done by Burke on a pinned beam illustrates the use of shaping polyvinylidene fluoride (PVDF) film, a piezoelectric material, in terms of singularity functions to design a sensor that is sensitive to the beam's velocity or moment at a specific location.¹ Burke demonstrated how a sensor with a given width at one end of the beam that decreases linearly to zero at the other end is sensitive to the beam's moment at the tip with the nonzero film width. Burke has also extended this work to two-dimensional systems.² Weighting functions that produce sensors that are sensitive to the modal vibration of a structure were demonstrated by Lee and Moon.³ Other shape functions and their benefits are discussed in Refs. 4–6. These works are significant in that they discuss the use of shape functions and windows to achieve frequency or wave number roll-off with distributed sensors.

Discrete sensors can also be used to obtain wave number information. Maillard and Fuller used an array of accelerometers and digital signal processing to achieve the desired wave-number filtering.⁷ This is accomplished by passing the accelerometer information through an array of finite impulse response filters that process the data to provide information on specific wave-number components. The wave-number filtering is achieved through software, as opposed to distributed sensors.

This paper presents results of a numerical investigation using an array of shaped PVDF sensors to act as low-pass

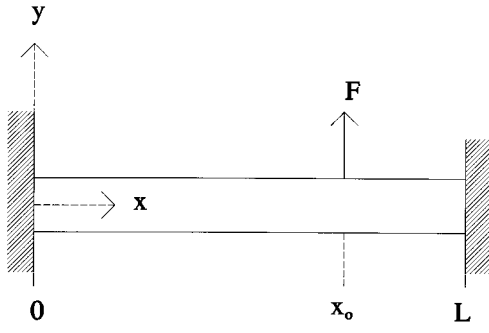


FIG. 1. Schematic of beam geometry for the system.

filters of the spatial strain transform. An analytical model of a clamped-clamped Bernoulli–Euler beam is presented to obtain the beam’s strain response, which is then used to obtain the response of point and distributed sensors. The choice of the distributed sensor shape is also investigated. Finally, the radiated power estimates using point and distributed sensors are compared to those predicted by theory.

I. ACOUSTIC RADIATION FROM STRUCTURES

The one-sided far-field acoustic power per unit width radiated from a beam undergoing sinusoidal vibration can be expressed as

$$\Pi = \frac{\omega \rho_f}{4\pi} \int_{-k_f}^{k_f} \frac{|V(k_x)|^2}{\sqrt{k_f^2 - k_x^2}} dk_x, \quad (1)$$

where ω is the angular frequency, ρ_f is the fluid density, $V(k_x)$ is the spatial Fourier transform of the beam velocity, k_x is the beam wave number, and k_f is the total fluid wave number.⁸ Equation (1) shows that only the wave number components with magnitude less than or equal to k_f , which are referred to as supersonic wave numbers, contribute to the far-field sound radiation. Therefore, if a suitable velocity wave-number spectrum is obtained, an estimate of the radiated power from the structure can be made using Eq. (1). Since the strain in the beam is related to the displacement, and hence the velocity, the strain wave-number transform can also be used to estimate the radiated power of the beam.

A. System model

The system studied is a Bernoulli–Euler damped beam of length L with clamped ends (see Fig. 1). The equation of motion for this system can be expressed as

$$E^* I \frac{\partial^4 \xi(x, t)}{\partial x^4} + m \frac{\partial^2 \xi(x, t)}{\partial t^2} = F_0 \delta(x - x_0) e^{j\omega t}, \quad (2)$$

$$E^* = E(1 + j\eta), \quad I = bh^3/12,$$

where ξ is the displacement in the y direction, E is the Young’s modulus, I is the second moment of the cross-sectional area for a rectangular cross section, m is the mass per unit length of the beam, F_0 is the magnitude of the forcing function, $\delta(x)$ is the Dirac delta function, η is the damping loss factor, b is the width of the beam, and h is the beam

TABLE I. Beam parameters used in the theoretical model.

| Parameter | Symbol | Value |
|---------------------|--------|------------------------|
| Length | L | 0.914 m |
| Width | w | 0.0508 m |
| Height | h | 0.006 35 m |
| Young’s modulus | E | 71 GPa |
| Mass density | ρ | 2700 kg/m ³ |
| Damping loss factor | η | 0.05 |

thickness. The specific beam values used in the analytical model are presented in Table I.

B. Natural frequencies

The transcendental equation for the system, obtained by setting the forcing function to zero in Eq. (2) and applying the appropriate boundary conditions, is

$$\cosh(k_b L) \cos(k_b L) = 1, \quad (3)$$

$$k_b = \left(\frac{\rho \omega^2}{E^* \kappa^2} \right)^{1/4}, \quad \kappa = \frac{h}{\sqrt{12}},$$

where k_b is the beam bending wave number ρ is the mass density of the beam, and κ is the radius of gyration for a rectangular section.⁹ To obtain the undamped natural frequencies, the damping loss factor was set to zero in the transcendental equation. The first five roots for $k_b L$ and their corresponding frequencies are shown in Table II.

C. Forced solution

The equation of motion can also be solved to obtain an expression for the beam displacement response. Polyvinylidene fluoride (PVDF) generates a voltage signal that is proportional to the strain. Therefore, an analytical expression for strain is desired. The strain, ϵ , can be obtained from the displacement using the relationship

$$\begin{aligned} \epsilon(x, t) &= -y \frac{\partial^2 \xi(x, t)}{\partial x^2} = \begin{cases} \epsilon_1(x, t), & x = [0, x_1], \\ \epsilon_2(x, t), & x = [x_1, L], \end{cases} \\ \epsilon_1(x, t) &= -y k_b^2 \{ A_1 [\cosh(k_b x) + \cos(k_b x)] \\ &\quad + B_1 [\sinh(k_b x) + \sin(k_b x)] \} e^{j\omega t}, \\ \epsilon_2(x, t) &= -y k_b^2 \{ A_2 \cosh[k_b(x - x_1)] \\ &\quad + B_2 \sinh[k_b(x - x_1)] - C_2 \cos[k_b(x - x_1)] \\ &\quad - D_2 \sin[k_b(x - x_1)] \} e^{j\omega t}. \end{aligned} \quad (4)$$

Here, y is the distance from the neutral axis; the subscripts 1

TABLE II. Roots and frequencies of beam characteristic equation.

| Mode | $k_b L$ | Frequency (Hz) |
|------|---------|----------------|
| 1 | 4.73 | 40.07 |
| 2 | 7.85 | 110.45 |
| 3 | 11.00 | 216.52 |
| 4 | 14.14 | 357.92 |
| 5 | 17.28 | 534.67 |

and 2 refer to the beam sections before and after the excitation point, respectively; $x=0$ refers to the left-hand side of each section; and $x_1=x_0$ and $x_2=L-x_0$ refer to the right-hand side of sections one and two, respectively. The subscripted letters in Eqs. (4) are constants defined as

$$\begin{aligned} A_1 &= \frac{F}{E^* I k_b^3 V_8}, \quad B_1 = V_3 A_1, \quad A_2 = V_4 A_1, \\ B_2 &= V_7 A_1, \quad C_2 = V_5 A_1, \quad D_2 = V_6 A_1, \end{aligned} \quad (5)$$

and the V_i terms are constants given by

$$\begin{aligned} V_1 &= \frac{(\text{sh}_1 + \text{si}_1)(\text{sh}_2 \text{co}_2 - \text{ch}_2 \text{si}_2) - (\text{ch}_1 - \text{co}_1)(\text{sh}_2 \text{si}_2 - \text{ch}_2 \text{co}_2 - 1)}{2(1 - \text{ch}_2 \text{co}_2)}, \\ V_2 &= \frac{(\text{ch}_1 - \text{co}_1)(\text{sh}_2 \text{co}_2 - \text{ch}_2 \text{si}_2) - (\text{sh}_1 - \text{si}_1)(\text{sh}_2 \text{si}_2 - \text{ch}_2 \text{co}_2 - 1)}{2(1 - \text{ch}_2 \text{co}_2)}, \\ V_3 &= \frac{V_1 - \text{ch}_1}{\text{sh}_1 - V_2}, \quad V_4 = V_1 + V_2 V_3, \quad V_5 = \text{ch}_1 - \text{co}_1 + (\text{sh}_1 - \text{si}_1)V_3 - V_4, \\ V_6 &= \frac{V_4 + (\text{sh}_2 \text{si}_2 - \text{ch}_2 \text{co}_2)V_5}{\text{sh}_2 \text{co}_2 - \text{ch}_2 \text{si}_2}, \quad V_7 = -\frac{V_4 \text{ch}_2 + V_5 \text{co}_2 + V_6 \text{si}_2}{\text{sh}_2}, \quad V_8 = -\text{sh}_1 + \text{si}_1 - (\text{ch}_1 + \text{co}_1)V_3 - V_6 + V_7. \end{aligned} \quad (6)$$

For compactness, Eqs. (6) and subsequent equations use the following notation to express frequently occurring trigonometric and hyperbolic constants:

$$\text{si}_i = \sin(k_b x_i), \quad \text{co}_i = \cos(k_b x_i), \quad \text{sh}_i = \sinh(k_b x_i), \quad \text{ch}_i = \cosh(k_b x_i), \quad (7)$$

where the index, i , can take the value 1 or 2. The spatial Fourier transform of the strain, $E(k_x, t)$, is given by

$$\begin{aligned} E(k_x, t) &= \int_0^L \epsilon(x, t) e^{-jk_x x} dx = -y k_b^2 \left\{ \frac{A_1}{k_b^2 + k_x^2} [(k_b \text{sh}_1 + j k_x \text{ch}_1) e^{-jk_x x_1} - j k_x] - \frac{A_1}{k_b^2 - k_x^2} [(-k_b \text{si}_1 + j k_x \text{co}_1) e^{-jk_x x_1} - j k_x] \right. \\ &\quad + \frac{B_1}{k_b^2 + k_x^2} [(k_b \text{ch}_1 + j k_x \text{sh}_1) e^{-jk_x x_1} - k_b] - \frac{B_1}{k_b^2 - k_x^2} [(k_b \text{co}_1 + j k_x \text{si}_1) e^{-jk_x x_1} - k_b] \\ &\quad + \frac{A_2}{k_b^2 + k_x^2} [(k_b \text{sh}_2 + j k_x \text{ch}_2) e^{-jk_x L} - j k_x e^{-jk_x x_1}] + \frac{B_2}{k_b^2 - k_x^2} [(k_b \text{ch}_2 + j k_x \text{sh}_2) e^{-jk_x L} \\ &\quad - k_b e^{-jk_x x_1}] - \frac{C_2}{k_b^2 - k_x^2} [(k_b \text{si}_2 - j k_x \text{co}_2) e^{-jk_x L} + j k_x e^{-jk_x x_1}] \\ &\quad \left. - \frac{D_2}{k_b^2 - k_x^2} [(k_b \text{co}_2 + j k_x \text{si}_2) e^{-jk_x L} + k_b e^{-jk_x x_1}] \right\} e^{j\omega t}. \end{aligned} \quad (8)$$

A typical strain response and strain transform spectrum is shown in Fig. 2 for a unit force located at $x_0=0.64$ m.

II. POWER IN TERMS OF THE STRAIN TRANSFORM

Given the strain transform, it is desirable to obtain an expression for radiated power in terms of this quantity. Making use of the relationship between the Fourier transform of a function, $F(k_x)$, and the Fourier transform of the n th derivative of this function, $f^{(n)}(x)$,¹⁰ namely

$$\mathcal{F}\{f^{(n)}(x)\} = (j k_x)^n F(k_x), \quad (9)$$

the displacement and strain transforms are related by the equation

$$E(k_x) = y k_x^2 D(k_x). \quad (10)$$

Using Eq. (10) and assuming $e^{j\omega t}$ time dependence, the power can be written in terms of the strain transform. This equation is

$$\Pi = \frac{\omega^3 \rho_f}{4 \pi y^2} \int_{-k_f}^{k_f} \frac{|E(k_x)|^2}{k_x^4 \sqrt{k_f^2 - k_x^2}} dk_x. \quad (11)$$

III. POINT SENSORS VERSUS DISTRIBUTED SENSORS

Individual strain sensors must be used to obtain an estimate of the beam wave-number spectrum. If point sensors are used, a sufficiently large number of sensors must be used to resolve the highest wave-number components of the wave-number spectrum. This number is based on the Nyquist criterion and if it is not met, then aliasing problems occur.

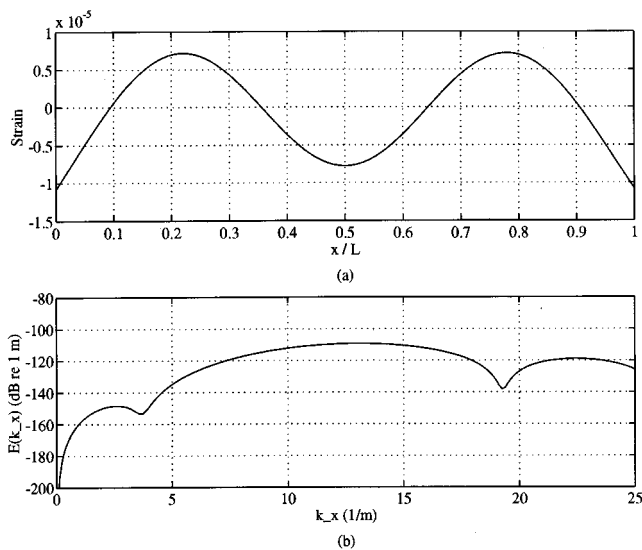


FIG. 2. Theoretical (a) strain and (b) Fourier transform of strain for resonance 3.

The Nyquist criterion specifies that a minimum of two sensors per wavelength must be used for the shortest wavelength that exists.

The number of sensors required to obtain a wave-number spectrum estimate without aliasing problems can be reduced by the use of distributed sensors. Distributed sensors can be fabricated to reduce the sensitivity of the sensors to higher wave-number components. In this way, the aliasing is

reduced and the number of sensors used can potentially be reduced.

A distributed strain sensor placed on the beam will output an electric signal proportional to the integrated strain over the area of the sensor. In equation form this proportion can be written as

$$\epsilon_d(x_s, t) = \int_{x_s-a}^{x_s+b} \epsilon(x, t) s(x) dx, \quad (12)$$

where ϵ_d is the distributed strain result; x_s is the center position of the sensor, $x_s - a$ and $x_s + b$ correspond to the minimum and maximum locations of the patch, respectively, and $s(x)$ is an arbitrary shape function. The form of the shape function used in this research is discussed in the next section.

IV. THE SHAPE FUNCTION

A distributed strain sensor effectively integrates the strain over the sensor area and a shape function can be introduced to vary the width of the sensor. This shape function can be chosen to provide spatial filtering. For this paper the shape function was chosen to provide low-pass filtering in the wave-number domain and is defined as

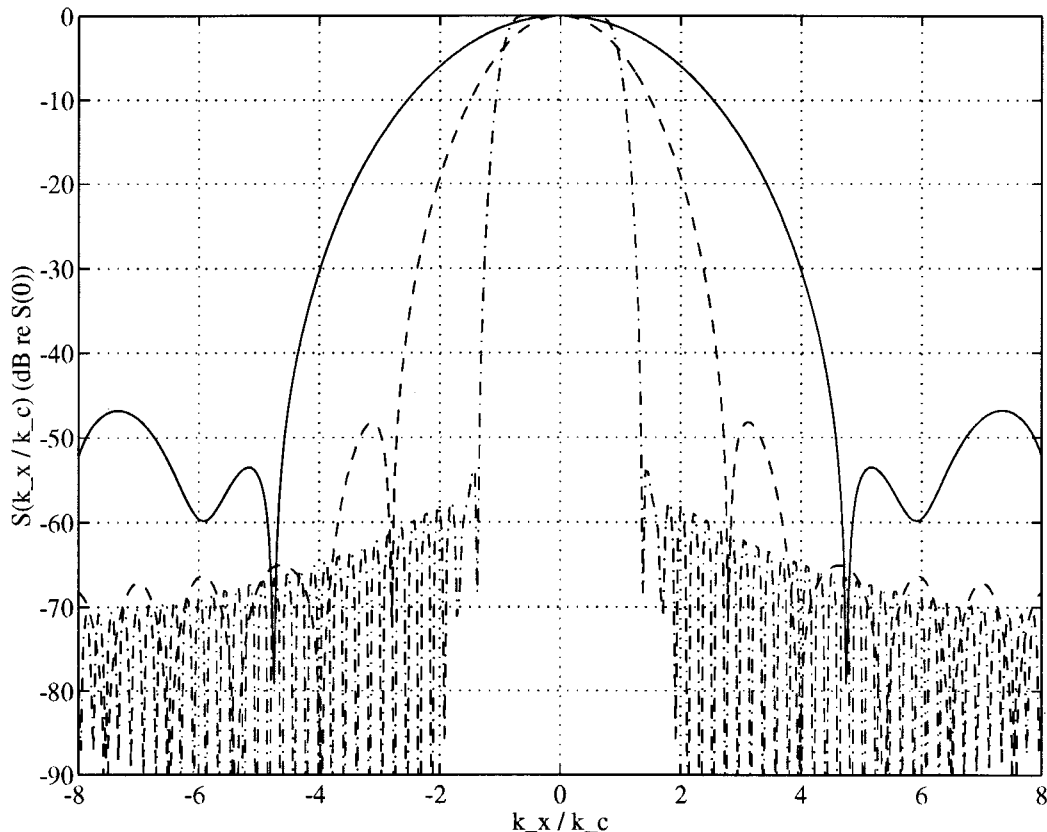


FIG. 3. Wave-number response of shape function. ($k_c l_p$) values shown are — 3, -- 6, and -·- 30.

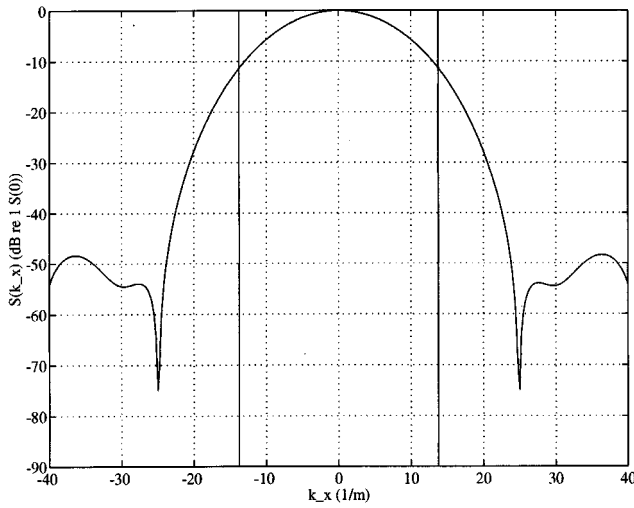


FIG. 4. Wave-number response of shape function for $l_p = 0.6$ m and $k_c = 6$ m⁻¹. The solid vertical lines indicate the Nyquist wave numbers.

$$s(x) = h(x - x_s) \text{sinc}[k_c(x - x_s)],$$

$$h(x) = \begin{cases} 0.54 + 0.46 \cos\left(2\pi \frac{x}{l_p}\right), & -a \leq x \leq b, \\ 0, & x < -a, \quad x > b, \end{cases} \quad (13)$$

$$\text{sinc}(x) = \frac{\sin(x)}{x},$$

where $h(x)$ defines a Hamming window,¹¹ k_c is the positive cutoff wave number, and l_p is the full sensor length. Note that l_p is the Hamming window length determined to provide the desired wave-number filter characteristics and is not necessarily equal to $a + b$. For example, if $s(x, x < 0)$ is not part of the structure and $s(x, x \geq 0)$ is part of the structure, then $a = 0$ and $b = l_p/2$. The corresponding Fourier transform of the shape function is

$$S(k_x) = \int_{-a}^b s(x) e^{-jk_x x} dx. \quad (14)$$

The sinc function provides the low-pass filter characteristics for the wave-number spectrum. A nontruncated sinc function would provide perfect low-pass wave-number filtering of the signal, while a finite function length causes the resulting signal to be nonzero for the domain $|k_x| > k_c$. The Hamming window was chosen to provide a smooth roll-off of the signal to minimize leakage when the discrete Fourier transform is used.

A closed-form solution for the integral of Eq. (14) is not possible. However, by making use of the convolution prop-

TABLE III. Location of point and patch sensors on the beam.

| Patch number | $x_s - a$ (m) | x_s (m) | $x_s + b$ (m) |
|--------------|---------------|-----------|---------------|
| 1 | 0 | -0.114 | 0.186 |
| 2 | 0 | 0.114 | 0.414 |
| 3 | 0.043 | 0.343 | 0.643 |
| 4 | 0.271 | 0.571 | 0.871 |
| 5 | 0.500 | 0.800 | 0.914 |
| 6 | 0.728 | 1.028 | 0.914 |

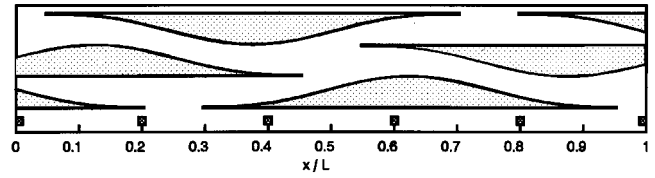


FIG. 5. Schematic showing patch and point sensor locations on the beam. Squares indicate point sensors.

erties of the Fourier transform, the perfect low-pass property of the sinc function and the finite window length $S(k_x)$ can be written in terms of the Fourier transform of the Hamming window. The resulting expression is given by

$$S(k_x) = \frac{e^{-jk_c x_s k_x}}{k_c} \int_{k_x - k_c}^{k_x + k_c} e^{j(k_c - 1)x_s \beta} H(\beta) d\beta, \quad (15)$$

where $H(k_x)$ is the Fourier transform of the Hamming window. For the special case of $x_s = 0$, Eq. (15) reduces to

$$S(k_x) = \frac{1}{k_c} \int_{k_x - k_c}^{k_x + k_c} H(\beta) d\beta. \quad (16)$$

Figure 3 plots the function $S(k_x)$ in Eq. (16) for various values of $k_c l_p$ and $a = b = l_p/2$.

V. NUMERICAL MODEL AND SIMULATIONS

The desire is to design an array of distributed sensors that can be used to estimate a beam's radiated power over a given frequency range. Since the array is intended to be used over a range of frequencies, the tuning of the system to a specific frequency or set of frequencies is not an acceptable solution. The three design parameters considered are the number of sensors, the full patch length, and the cutoff wave number.

A. Description of model and parameters

The beam system is identical to that outlined in Sec. I. The amplitude of the forcing function is set to unity and the forcing location is 0.64 m (0.7 L) for all runs. The frequency

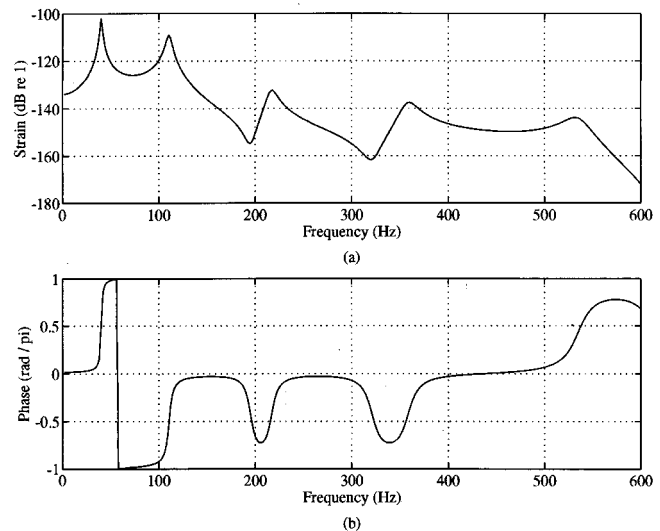


FIG. 6. Numerical frequency response for patch 3. (a) Magnitude and (b) phase.

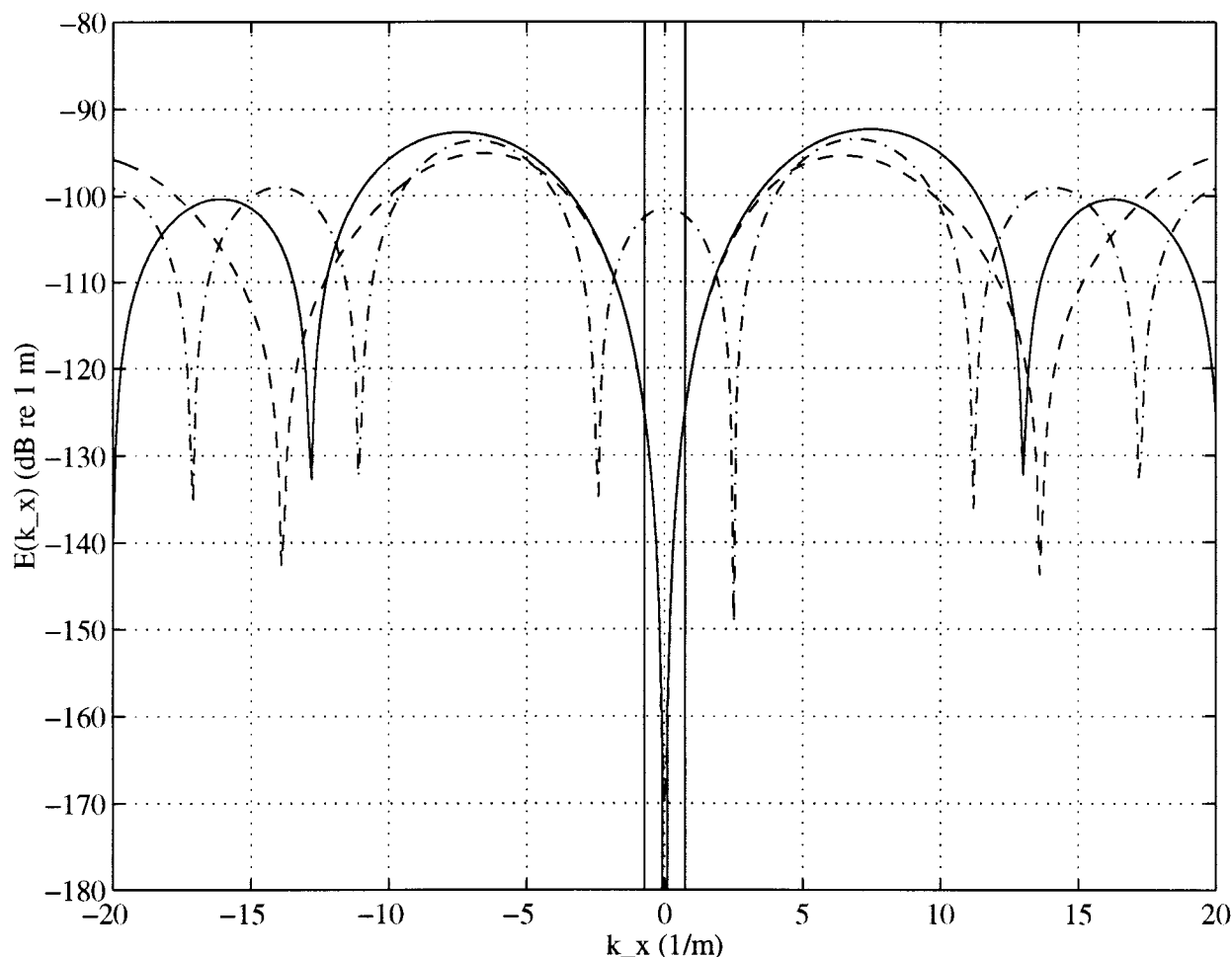


FIG. 7. Fourier transform of strain response for resonance 1 (40.07 Hz). Legend: — theoretical response, -- patch response, --- point response, and | acoustic wave number.

of the forcing function changes for the various tests carried out. To determine the number of sensors required, one must first calculate the center to center sensor spacing Δx , which is dependent upon the highest fluid wave number to be included in the analysis. The application of interest in this paper is active control of radiation from structures, and is thus concerned primarily with the low-frequency behavior of the beam. Therefore, analysis will be limited to radiation into air for frequencies only up to the fifth beam resonance frequency.

Using a fluid wave speed of 343 m/s and a value of 535 Hz for the fifth resonance frequency, the highest wave number of interest is 9.8 m^{-1} . The maximum sensor spacing allowed to sufficiently sample up to this wave number, as set forth by the Nyquist criterion, is 0.32 m. Using this sensor spacing and requiring that the first and last sensors whose center positions, x_s , are located on the beam be placed $\Delta x/2$ m from the beam ends, the minimum number of sensors required is three. A margin of safety is added by specifying that there be at least four sensors with center positions on the beam. The resulting sensor spacing is 0.23 m, which can resolve wave numbers up to 13.75 m^{-1} .

The next parameters to be calculated are the patch length and cutoff wave number, which define the shape function.

These two parameters are chosen to attenuate the wave-number components above the highest wave number resolvable as calculated above. Ideally, the sensors would provide very large attenuation above the design cutoff wave number, such as that provided by the $(k_c l_p) = 30$ curve in Fig. 3. However, if a cutoff wave number of 9.8 m^{-1} is chosen, the corresponding patch length of 3.06 m becomes excessively large for the beam. Alternatively, if a more reasonable patch length of 0.5 m is chosen, then the corresponding cutoff wave number of 60 m^{-1} becomes excessively large for the sensor spacing chosen.

A compromise between the patch length and cutoff wave number must be made. A patch length of 0.6 m and a cutoff wave number of 6 m^{-1} provides a reasonable compromise. This combination of parameters provides attenuation of approximately 6 dB at 9.8 m^{-1} , approximately 11 dB at 13.75 m^{-1} , and 20 dB or more for wave numbers above 18 m^{-1} (see Fig. 4). Although these attenuation levels are not as high as one might desire, the length and cutoff wave-number values chosen provide sufficient attenuation of higher wave numbers to obtain reasonable estimates of the spectrum over the frequency and wave-number range of interest.

To estimate the radiated power, the frequency domain

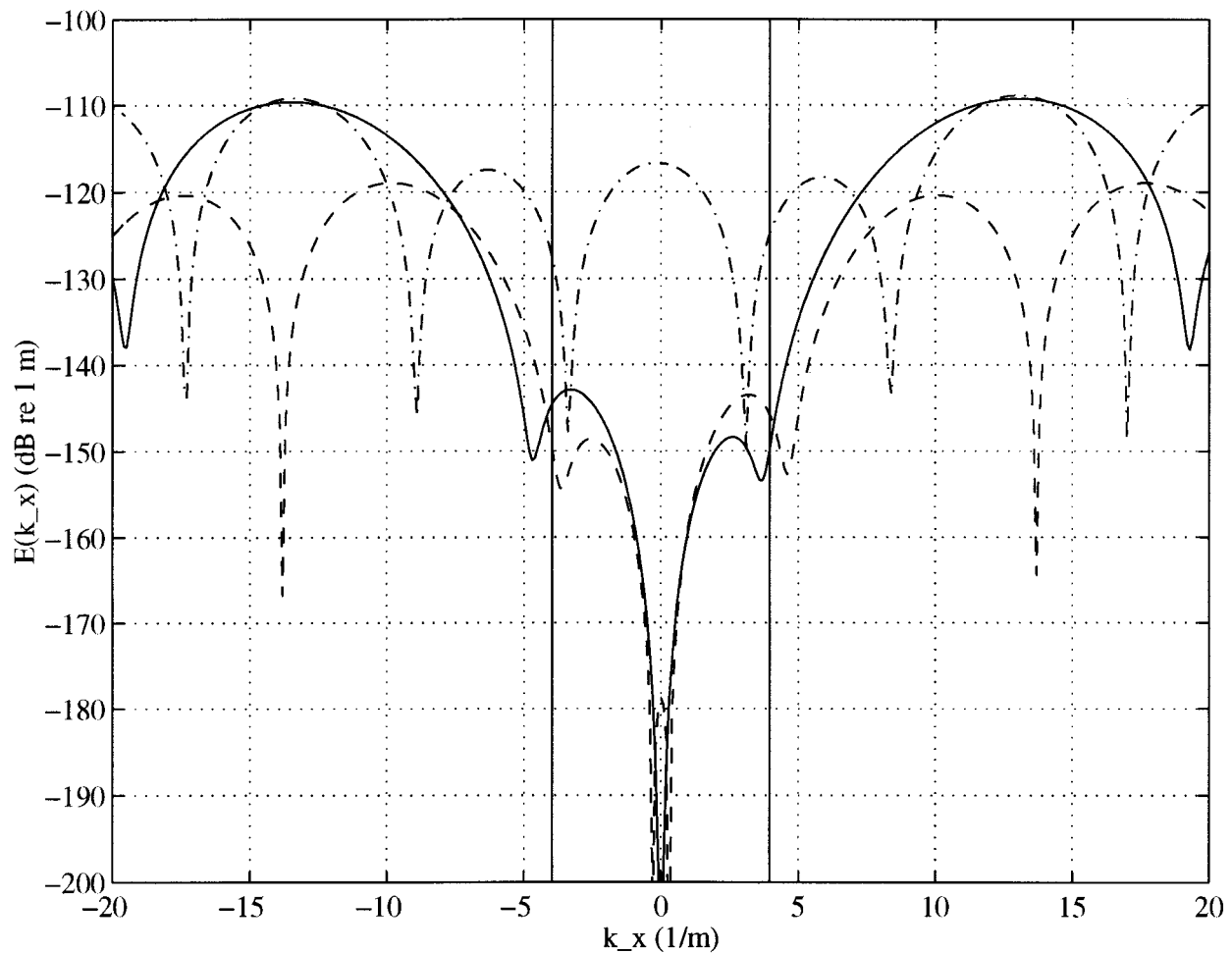


FIG. 8. Fourier transform of strain response for resonance 3 (216.52 Hz). Legend: — theoretical response, -- patch response, -.- point response, and . . . acoustic wave number.

response of each patch is obtained from the time-domain signal using a temporal Fourier transform. A spatial discrete Fourier transform (DFT) of the array of frequency domain signals then provides an estimate of the wave-number spectrum. Given the excitation frequency, the fluid wave number can be determined. This fluid wave number and the wave-number spectrum can then be used to estimate the radiated power by implementing Eq. (11).

The full patch length of 0.6 m means that there are six full or partial patches on the beam. Table III provides the center and end-point locations for each patch and Fig. 5 shows a schematic of the patch locations on the beam.

For comparison purposes, point strain sensors on the beam are also modeled. Six equally spaced point sensors were placed along the length of the beam. The locations of these sensors on the beam are 0, 0.183, 0.366, 0.548, 0.731, and 0.914 m (see Fig. 5). Note that no consideration was given to the optimal placement of the point sensors.

B. Numerical results

Two types of numerical tests were carried out. The first is a frequency response for each patch. This was accomplished by sweeping the driving frequency of the force over

the desired frequency range and recording the response level of each patch. The frequency response of patch 3 is shown in Fig. 6.

The second type of test is a spatial DFT of the six patches at a given frequency, which corresponds to the strain wave-number transform. First the model was driven at 40.07 Hz, which corresponds to the first beam resonance frequency, and the patch levels recorded. A DFT of the data was then taken to obtain a strain transform spectrum. This procedure was repeated for the frequencies 110.45, 216.52, 357.92, and 534.67 Hz, which correspond to the next four resonance frequencies of the beam. In addition to the patch strain transforms, the theoretical and point strain transforms were also calculated for these frequencies. Sample plots of these results are shown in Figs. 7–9. Next, the excitation frequency was set to 75, 160, 300, and 450 Hz, which are frequencies approximately midway between the resonance frequencies, and again a DFT of the strain results was calculated. Samples of these results are shown in Figs. 10 and 11.

Two items must be noted concerning these plots. First, to offset the attenuation of the strain field due to the shape function, the shape function was normalized so that the integral of the shape function for a full patch was unity,¹² i.e.,

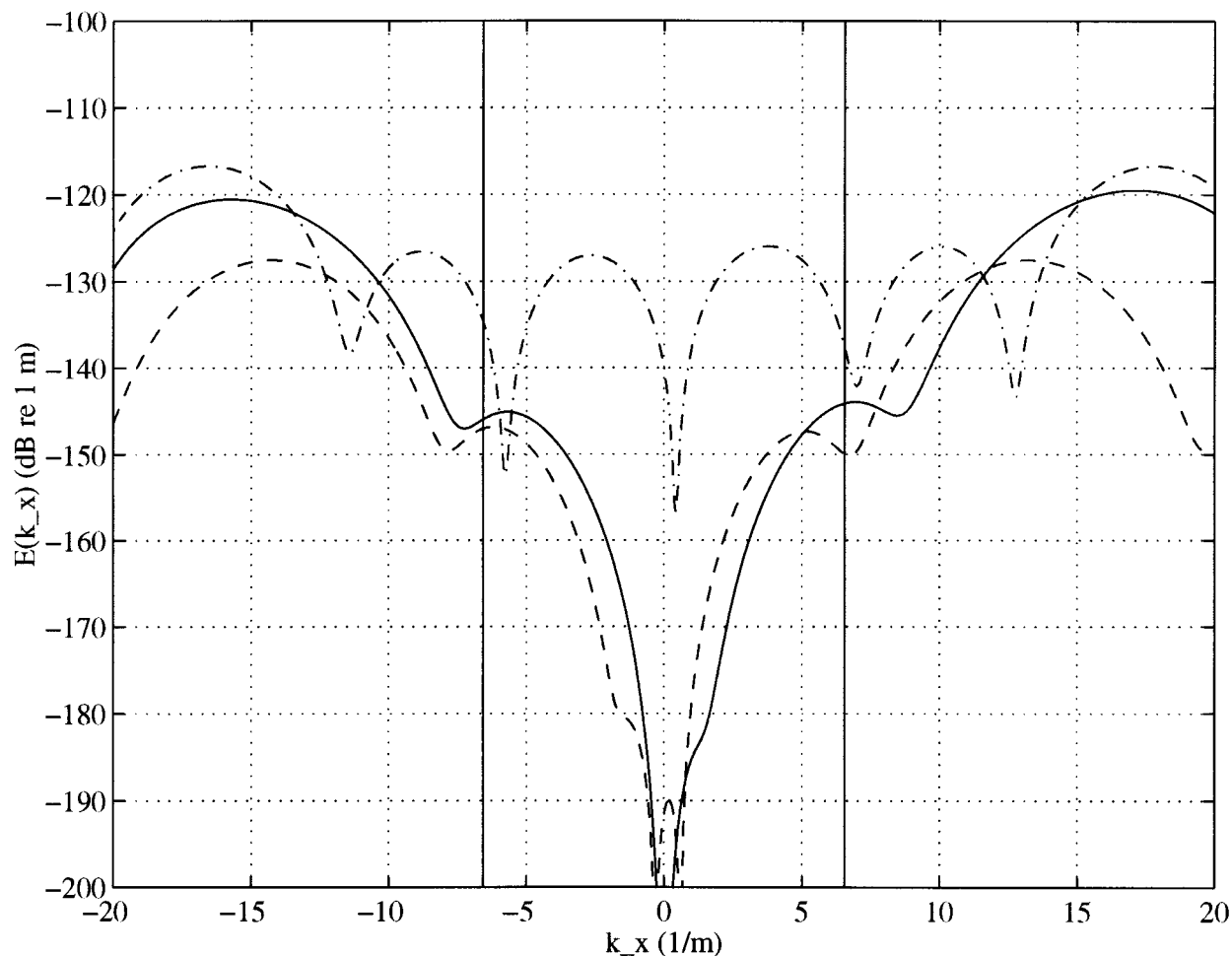


FIG. 9. Fourier transform of strain response for resonance 4 (357.92 Hz). Legend: ——— theoretical response, --- patch response, -.- point response, and $|k_x|$ acoustic wave number.

$$\int_{-0.3}^{0.3} s(x) dx = 1. \quad (17)$$

The application of this normalization introduces a shift of 10.5 dB to the filtered strain transforms. Second, to account for aliasing associated with the point sensors, the point strain transforms were shifted by -5.4 dB. This number represents the average overestimation of the point sensors at the first four resonance and first two off-resonance frequencies when compared to the maximum level of the theoretical strain transforms. The use of this correction for the point spectra assumes that the maximum level of the point spectrum should match the maximum level of the theoretical spectrum if the peak of the theoretical spectrum lies below the point Nyquist wave number.

The final computations involve calculating the integral of $|E(k_x)|^2$ over the range of supersonic wave numbers and estimating the radiated power from the beam using the strain transform results. Specific strain integrals were evaluated at the first five resonance frequencies and four frequencies between the resonance frequencies. Power estimates were obtained by performing a frequency sweep and implementing Eq. (11). The integral calculations use Simpson's 1/3 rule¹³ to approximate the strain transform integral and the integral in Eq. (11). To circumvent problems associated with the

poles at $-k_f$, 0 , and k_f of the radiated power equation, the range of integration used was $-0.95k_f \leq k_x \leq -0.05k_f$ and $0.05k_f \leq k_x \leq 0.95k_f$. Each wave-number range was then divided into 50 equally spaced intervals and the theoretical, patch, and point transforms calculated. Shift factors were applied to the point and distributed strain transforms, as explained in the preceding paragraph. The results of the point and patch strain integral are presented in Table IV. The power predictions are presented in Fig. 12.

C. Discussion

The frequency response for patch 3 (Fig. 6) clearly shows the expected first five resonances of the beam.

The strain transforms (Figs. 7–11) show that the patch results follow the theoretical values reasonably well for structural wave numbers below the fluid wave number, while the point results show significant differences from the theoretical results, not only in level, but also in form. Examining the predicted response at the first resonance frequency (Fig. 7) shows a number of key elements. The theoretical values go to zero at low wave numbers. This is an expected result, given the relationship between the strain and displacement

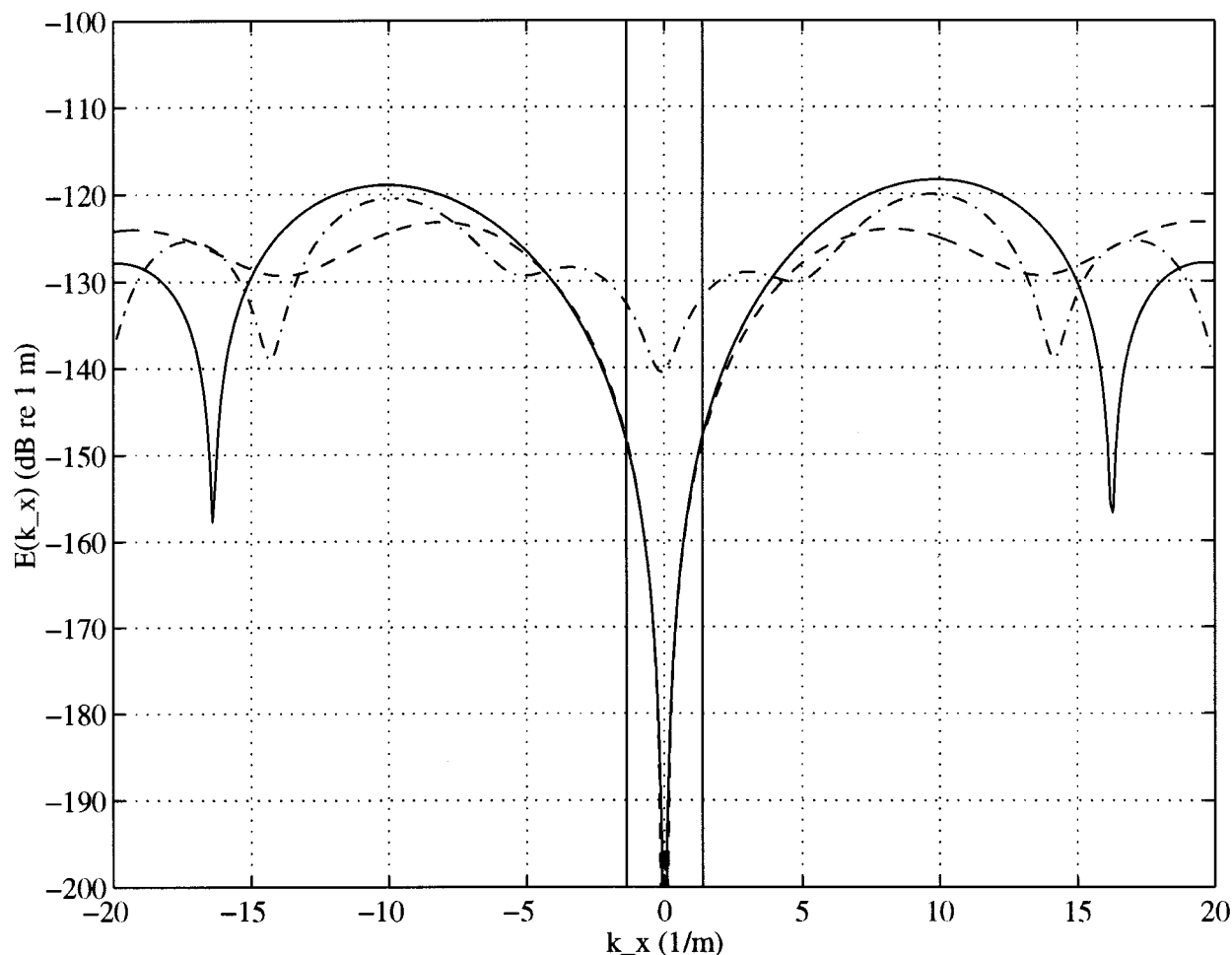


FIG. 10. Fourier transform of strain response for 75 Hz. Legend: — theoretical response, -- patch response, ··· point response, and | acoustic wave number.

spatial transforms [see Eq. (10)] and is expected for all frequencies. The patch results show excellent agreement with the theoretical results. The patch result tends towards zero at low wave numbers with a level of approximately -170 dB at $k_x = 0$. Conversely, the point results differ sharply from the theoretical values. The point result has a local maximum level near zero wave number and does not follow the shape of the analytical solution below the fluid wave number. This high level at low wave number is due to aliasing problems associated with the point sensors.

Given the sensor spacing and using the Nyquist criterion, the highest wave numbers that the point and patch arrays can resolve are 17.19 and 13.75 m^{-1} , respectively. As can be seen from the theoretical curve, the strain transform levels are high above these k_{max} values. Unlike the patches, which provide wave-number filtering to reduce the level of the high wave-number components, the point sensors provide no filtering and all of the wave-number components above k_{max} are aliased back to the lower spectral lines. Another aspect that is apparent in Fig. 7 is the wrapping characteristic associated with the DFT. Both the point and patch results show a symmetry about their respective k_{max} values.

Looking at the results of the third resonance (Fig. 8), one sees that the patch array has a similar shape to the the-

oretical curve. However, the locations and levels of their maxima do not correspond. The patch response overestimates the theoretical value by approximately 6 dB for positive wave numbers and underestimates the theoretical level by a similar amount for negative wave numbers. A deviation at low wave number is noticeable on the patch curve. However, this deviation occurs in the vicinity of -180 dB and does not have a significant effect on the overall shape of the curve. The point curve does not follow the shape of the theoretical curve and again has a high response at low wave numbers.

The response at the fourth resonance (Fig. 9) also shows a marked difference between the theory and patch levels. The levels of the patch response are higher than the theoretical values from a wave number of approximately 0.6 – 5 m^{-1} and lower than the theoretical values from 5 m^{-1} up to the maximum wave number of interest, which is the fluid wave number at 6.56 m^{-1} . The patch level falls below the theoretical level for most of the negative wave-number range of interest. The point levels are again significantly different from the theoretical values, with the maximums occurring well before the theoretical values.

The off-resonance response plots show similar results to the resonance plots. That is, the patch results follow the the-

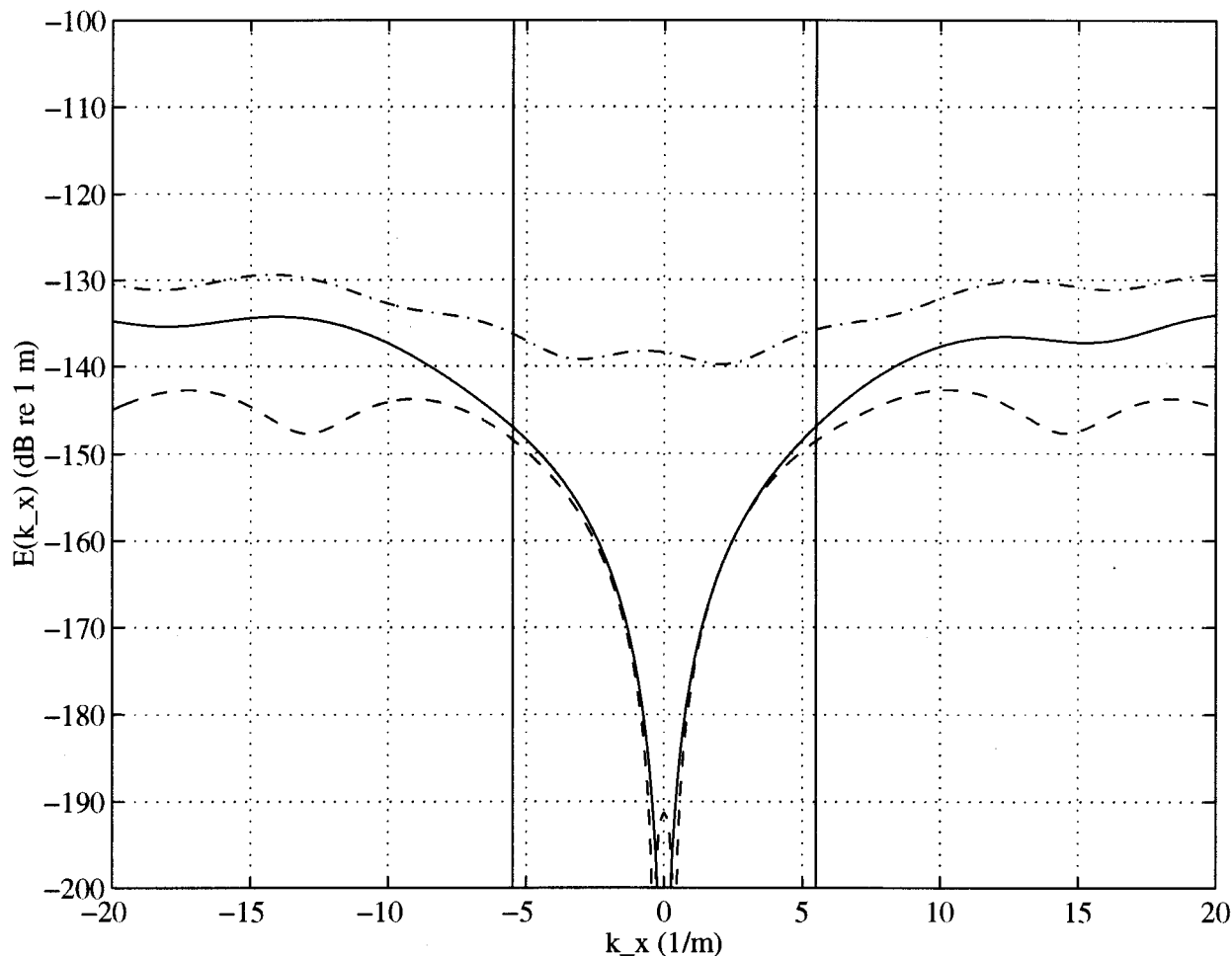


FIG. 11. Fourier transform of strain response for 300 Hz. Legend: ——— theoretical response, --- patch response, -.- point response, and | acoustic wave number.

oretical curve below the fluid wave number while the point response is affected by aliasing. The aliasing problem for the point array seems to be apparent at each driving frequency and could be due to the mixture of multiple modes of the beam. As before, the low wave-number deviations in the patch response become more pronounced at higher frequencies, but are still well below the maximum levels.

The values for the integrals of the squared magnitude of the strain transform in Table IV show that the patch results

range from -1.6 to 3.0 dB when compared to the theoretical results. These deviations from the theoretical results, which are the integral estimates for resonances four and five, are not surprising considering the differences in the strain transforms predicted by theory and the patch array. Predictions for the other frequencies studied are all within 1 dB of the theoretical result. Unlike the patch results, the integral using the point sensors overestimated the theoretical values by 15.8 – 30.9 dB. These high overestimation predictions are caused by aliasing problems associated with the point array.

Since the ultimate goal of this research is to determine if shaped sensors can provide a means for obtaining an acceptable estimation of the radiated power, the power results presented in Fig. 12 provide an excellent criterion for assessing the performance of the point and patch sensors. As with the results for the integrals of the squared magnitude, the power calculations show the power calculations using the patch sensors are significantly better than the power results using the point sensors, when compared to the theoretical values. Both the point and patch results show a marked increase in predicted radiated power at low frequencies. However, if only the frequency range between 20 and 600 Hz is consid-

TABLE IV. Integral of magnitude squared strain transform for patch and point arrays.

| Frequency (Hz) | $\int E(k_x) ^2 dk_x$ (dB re: theory) | |
|----------------|--|---------------|
| | Patch sensors | Point sensors |
| 40.07 | -0.1 | 30.9 |
| 110.45 | -0.4 | 26.5 |
| 216.52 | -0.4 | 27.6 |
| 357.92 | -1.6 | 20.9 |
| 534.67 | 3.0 | 19.5 |
| 75 | -0.1 | 20.2 |
| 160 | -0.4 | 20.6 |
| 300 | -1.0 | 15.8 |
| 450 | 0.7 | 19.5 |

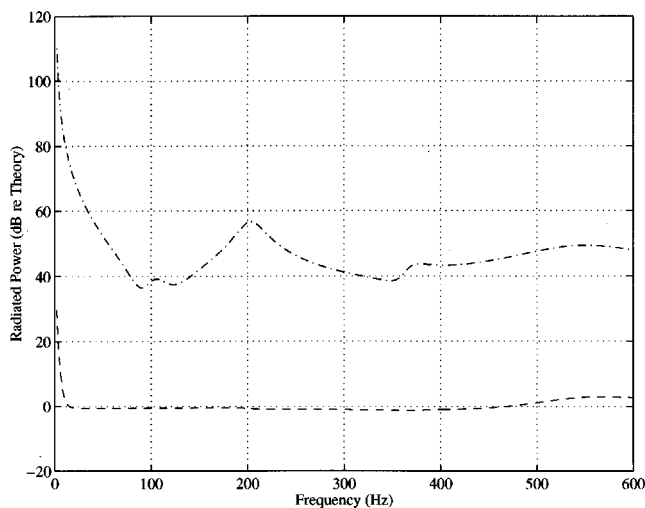


FIG. 12. Estimated power of patch and point arrays. Legend: — patch response and --- point response.

ered, the spread for the patch sensor estimates is -1 to 3 dB, while the point sensor estimates have a spread of 36 – 68 dB.

VI. CONCLUSIONS

This paper presents a technique to obtain the far-field radiated power from a one-dimensional structure using an array of shaped polyvinylidene fluoride sensors. The distributed sensors are shaped to act as low-pass wave-number filters, since the radiated power depends only on the structural wave-number components whose magnitudes are less than or equal to the fluid wave number.

The results show that the shaped sensors provide a better approximation to the theoretical strain transform than do the point sensors. This is verified by the form of the functions, presented by plotting the individual strain transforms, and by integrating the squared magnitude of the strain transforms and comparing the results to the theoretical values. Overall, these results indicate good agreement with the analytical predictions. The errors for the integral of the squared magnitude of the strain transforms are all 3 dB or less when using the shaped sensors, compared to at least 15.8 dB when using the point sensors.

Power estimates using the strain transform results show that patch estimations are again considerably better than the point estimations when compared to the theoretical predictions. The power results indicate that the use of shaped sen-

sors, as described in this paper, can provide a better estimate of a structure's radiated power than the use of discrete point sensors. The estimated powers using the shaped sensors are all within 3 dB of the actual power for most of the frequency range, while those using discrete point sensors overestimate the level by at least 36 dB for the same frequency range. These results indicate the ability for shaped sensors to provide significantly improved power estimates for applications that estimate radiated power from structural measurements.

ACKNOWLEDGMENTS

This work was supported by the Applied Research Laboratory at The Pennsylvania State University and the Acoustical Society of America.

- ¹S. E. Burke, "Distributed transducer shading: Application to structural control, hydrodynamics, and sonar sensing," *Proc. SPIE* **1917**, 660–673 (1993).
- ²S. E. Burke and J. E. Hubbard, Jr., "Distributed transducer vibration control of thin plates," *J. Acoust. Soc. Am.* **90**, 937–944 (1991).
- ³C.-K. Lee and F. C. Moon, "Modal sensors/actuators," *J. Appl. Mech.* **57**, 434–441 (1990).
- ⁴S. A. Collins, D. W. Miller, and A. H. von Flotow, "Piezopolymer Spatial Filters for Active Structural Control," in *Proceedings of the Conference on Recent Advances in Active Control of Sound and Vibration*, edited by C. A. Rogers and C. R. Fuller (Technomic, Lancaster, PA, 1991), pp. 219–234.
- ⁵D. W. Miller, S. A. Collins, and S. P. Peltzman, "Development of Spatially Convolution Sensors for Structural Control Applications," Collection of Technical Papers—AIAA/ASME/ASCE/AHS Structures, Structural Dynamics & Materials Conference, Part 4 (American Institute of Aeronautics and Astronautics, Washington, DC, 1990), pp. 2283–2297.
- ⁶M. S. Anderson and E. F. Crawley, "Discrete Shaped Strain Sensors for Intelligent Structures," *Proceedings of the AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Material Conference* (American Institute of Aeronautics and Astronautics, Washington, DC, 1992), pp. 566–576.
- ⁷J. P. Maillard and C. R. Fuller, "Advanced time and wave-number sensing for structural acoustic systems. I. Theory and design," *J. Acoust. Soc. Am.* **95**, 3252–3261 (1994).
- ⁸F. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic, New York, 1985), p. 76.
- ⁹P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton U. P., Princeton, NJ, 1968), p. 185.
- ¹⁰M. D. Greenberg, *Foundations of Applied Mathematics* (Prentice-Hall, Englewood Cliffs, NJ), p. 102.
- ¹¹F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE* **66**, 51–83 (1978).
- ¹²M. S. Anderson and E. F. Crawley, "Discrete Shaped Strain Sensors for Intelligent Structures," *Proceedings of the AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Material Conference* (American Institute of Aeronautics and Astronautics, Washington, DC, 1990), p. 567.
- ¹³W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge U. P., New York, 1988), p. 114.

Comparison between subjective and objective measures of active hearing protector and communication headset attenuation

Jan Zera,^{a)} Anthony J. Brammer, and George J. Pan

National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

(Received 17 April 1996; revised 13 December 1996; accepted 29 December 1996)

A masked-threshold and a loudness-balance method have been developed to estimate the attenuation of communication headsets and hearing protectors with built-in active noise reduction (ANR) systems. Both methods are used to estimate the attenuation of the ANR systems and the masked-threshold method is also used to estimate the total attenuation (active plus passive) of the device. The procedures are designed to be used in the presence of environmental noise, and to minimize the noise exposure of subjects during the measurements. For comparison, physical measurements of insertion loss have also been performed using a miniature microphone in the concha. Experiments showed that the masked-threshold method tends to give increased estimates of the attenuation if the noise reduction of the left and right earcup ANR systems differs, as commonly occurs in practice. In contrast, the loudness-balance method reduces the estimates of the active attenuation. Insertion loss measurements may be influenced by the position of the microphone, owing to the spatial variability of the sound field under an earmuff when the ANR system is operating. Differences between physical and subjective measurements of up to 20 dB have been obtained in this study at frequencies of 250 Hz and below for a device in which the sound pressure varied substantially near, and within, the ear canal. [S0001-4966(97)02205-4]

PACS numbers: 43.50.Hg, 43.50.Ki [GAD]

INTRODUCTION

There are two commonly used designs for hearing protectors and headsets (protectors incorporating a communication channel) with built-in active noise reduction (ANR) systems. One employs a large circumaural earcup with a compliant cushion. In this device, the ANR system is used to increase the attenuation in the frequency range below 500 Hz. At mid and high frequencies, the headset displays considerable passive attenuation. The other type of device employs lightweight, often supra-aural construction. The ANR system of this headset provides most of the headset attenuation at frequencies from 50 to 1000 Hz. All these active devices differ from traditional passive devices in several ways that are important for the measuring methods used for their assessment. Firstly, the presence of external (environmental) noise is required for their proper operation. Secondly, the attenuation of the ANR system often intentionally changes with the external noise level. Thirdly, the electronic noise of the system may cause additional masking in real-ear-at-threshold (REAT) measurements,¹⁻³ leading to the overestimation of attenuation.^{4,5} Thus, an active hearing protector, or headset, cannot be tested at noise levels corresponding to the absolute threshold of hearing, nor can the device be assumed to be linear over a wide range of sound levels. These factors suggest the need for measurement methods specifically designed for devices with ANR systems.^{6,7,4,8}

There are, at present, few reports evaluating the performance of active hearing protectors or headsets using subjective

methods. Forshaw *et al.*⁹ used a masked-threshold and a loudness-balance method to assess the active attenuation of an ANR headset, and Crabtree and Rylands¹⁰ have also used a loudness-balance method for this purpose. Casali and Robinson⁴ have discussed the difficulty of establishing a uniform, standardized procedure for measuring the attenuation of all types of active hearing protectors, and proposed a hypothetical hybrid procedure suitable for devices in which the ANR is used only to increase the attenuation at low frequencies. According to their proposal, the standardized REAT method would be used for subjective measurement of the passive attenuation of the device, while objective measurement would be used to determine the total (active+passive) attenuation and the ANR.

We believe that subjective and objective methods should be developed independently for the measurement of total attenuation, which is the most important parameter to the user, as well as for the active component (ANR), which has to be determined during device development. Because an ANR headset or hearing protector should be tested at noise levels similar to those at which it will be used, suprathreshold psychoacoustic methods would appear to be required for this purpose.

The purpose of this research was to develop and test methods of measurement suitable for the ANR devices, in an attempt to foster the development of standardized methods of measurement. We present here two subjective methods for estimating the active attenuation of an ANR headset or hearing protector: a masked-threshold procedure and a loudness-balance procedure. The masked-threshold procedure may also be used to estimate the total attenuation, and is designed to minimize the subject's exposure to noise. The loudness-

^{a)}Current address: Central Institute for Labour Protection, ul. Czerniakowska 16, 00-701 Warsaw, Poland.

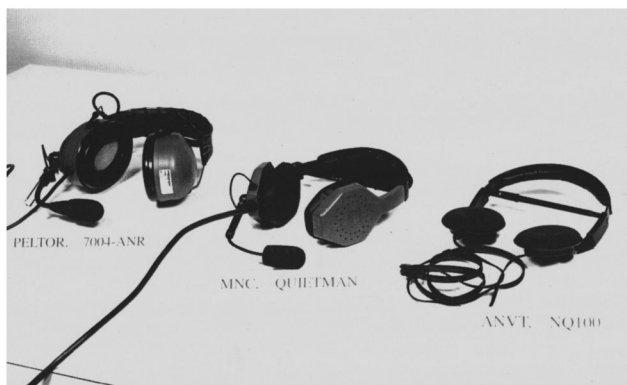


FIG. 1. Three devices used in the experiments.

balance procedure differs from its traditional form by employing a two-alternative forced-choice (2-AFC) design, and is implemented using an adaptive technique. This allows us to simplify the experiment for inexperienced listeners and so improve their decisions. The results of the subjective measurements are then compared with objective measurements of the attenuation recorded by a miniature microphone located either in the concha^{11,12} or within the ear canal.

In the following sections, we first introduce the ANR devices, measuring system, and subjects. Then, the subjective and objective measuring methods are described and results reviewed. This order was chosen since it allowed us more opportunity to discuss the measurement methods.

I. APPARATUS AND SUBJECTS

A. Hearing protector and headsets

Two communication headsets and one hearing protector, representing various design concepts, were selected for the experiments (see Fig. 1). These were a Peltor headset (model 7004), a QuietMan headset (manufactured by MNC, Inc.), and a hearing protector, model NQ100, manufactured by Active Noise and Vibration Technologies (ANVT).

The Peltor 7004 is a circumaural headset with a large volume (approximately 130 cm³) enclosed by the earcup (see Fig. 1, the sample to the left). A combined liquid-filled and foam cushion provides an effective seal between the earcup and the head. The ANR system of this headset is designed to improve its attenuation at frequencies below 300 Hz.

The QuietMan headset (see Fig. 1, the middle sample) is also a circumaural device, but with a smaller volume under the lightweight earcup and a partial air seal between the earcup and the head. The attenuation of this headset at frequencies below 1000 Hz is entirely dependent on the performance of the ANR system.

The ANVT NQ100 hearing protector (see Fig. 1, sample to the right) is a supra-aural device of very lightweight construction with the ANR system acting in a restricted frequency range (70–400 Hz), and with some passive attenuation above 3000 Hz.

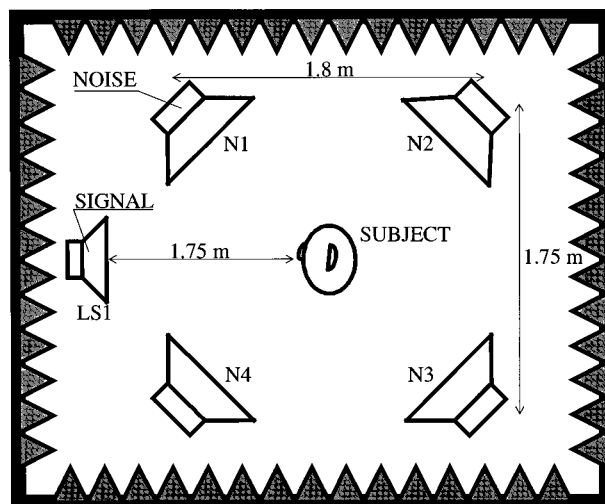


FIG. 2. Floor plan of the experimental setup in the anechoic chamber. N1–N4: four-speaker source of pink noise. LS1: speaker used to present test signal when the headset was not worn by the subject.

B. External noise source

Measurements were conducted in an anechoic chamber (inner dimensions 5.6×3.5×3.4 m) located on the campus of the National Research Council in Ottawa. A special noise source was constructed for the measurements, consisting of four loudspeaker systems located in a horizontal plane at the corners of a rectangle 1.8 m by 1.75 m (see Figs. 2 and 3). Each loudspeaker unit included a subwoofer (State of Art Elektronik, 300 W), a midrange section (PSB-800 100-W two-way loudspeaker system), and an Audax super tweeter. The Audax super tweeters were used in addition to the tweeters of the PSB speakers to enhance the performance of the system at high frequencies. A Bryston active crossover (model 10B) was used to divide the frequency range into low (20–100 Hz), mid (100–4500 Hz), and high (4500–20 000 Hz) frequency bands. The low- and midrange speakers were powered by four Bryston 4B amplifiers (250 W per channel). Bryston 2B amplifiers (50 W per channel) were used to drive the Audax super tweeters.

The noise source was capable of generating broadband (25–20 000 Hz) pink noise of 110 dB sound-pressure level



FIG. 3. Experimental setup in the anechoic chamber. Speaker LS1 is not shown.

(SPL) at a reference point corresponding to the center-head position with the subject and chair absent. Before experiments commenced, the noise level measured in one-third octave bands at the center-head position (with subject and chair absent) was adjusted electronically to be within ± 3 dB of the target pink noise spectrum, using a Technics Stereo Equalizer model SH-9010. If necessary, the positions of the high-frequency tweeters were also adjusted. The diffuseness of the sound field at the subject's head was increased by introducing a 25-ms delay between midfrequency speakers using a Klark-Teknik model DN700 delay line.

C. Subjects

Seven subjects, aged from 20 to 24 years, participated in the experiments. Before experiments commenced their hearing thresholds were measured using a clinical audiometer (Madsen, model OB 40) in accordance with ANSI S3.6-1989. Subjects had normal hearing, with less than 10 dB HL throughout the frequency range from 125 Hz to 8 kHz. Five subjects (S1–S5) participated in both the initial psychoacoustical tests and the objective measurements described in Secs. II and III. Two subjects (S6 and S7) participated in measurements of the spatial variability of the sound field under the earcup described in Sec. IV. The number of subjects was chosen so that, within the time constraints available for the study, measurements could be repeated for each set of experimental conditions.

A subject, wearing an ANR device, was seated in a chair at the center of the four-speaker source (see Fig. 2). The subject's head position was referenced in the horizontal plane to the midpoint of the loudspeaker array, which was defined by the location of the tweeters of the PSB speakers. The reference point was maintained by two (horizontal and vertical) laser pointers, with an accuracy of ± 0.5 cm. The chair was equipped with a support to position the back of the subject's head during measurements (Fig. 3). For safety reasons, and to improve communication between the subject and the experimenter, a video camera was mounted in the anechoic chamber. It allowed the subject to be seen by the experimenter, who continuously monitored the subject's head position during the measurement session. A measurement session, involving one headset or the hearing protector, was completed in 3 h.

The subjects had no prior experience of psychoacoustical measurements and were initially trained for about 5–10 h until learning effects were not observed.

II. SUBJECTIVE METHODS

A. Masked-threshold method

In the masked-threshold method, the difference in threshold between that determined in a noisy environment when the headset, or hearing protector, is worn and when it is not worn provides a measure of the attenuation of the device.^{13,11} The complication associated with using an additional sound source may explain why the masked-threshold method has never been standardized, and is infrequently used to evaluate passive hearing protectors. Active hearing protectors and headsets are already equipped with an earphone to

serve as the secondary source for active noise cancellation. In headsets, and some active hearing protectors, the earphone is also used to deliver signals from a communication channel. In these circumstances, it is feasible to use this channel to present the test signals for a masking procedure without compromising the performance of the ANR system.

When a conventional masked-threshold method is employed, masking noise of up to 100 dB SPL may be required in order to measure the device's performance under realistic operating conditions. A health and ethical problem then arises as to how to protect subjects from excessive exposure to high noise levels, and to prevent temporary threshold shift from influencing the results. White noise, or bandlimited white noise, has been used in most masking experiments.¹⁴ However, the shape of the hearing damage risk contours as a function of frequency¹⁵ suggests that white noise would produce either insufficient sound levels at low frequencies, or excessive sound levels at high frequencies. For the purpose of the present experiments, pink noise would appear to provide a reasonable balance between energy at high and low frequencies, and was therefore adopted for the masking noise. During the experiments the level of broadband pink noise was set to a maximum of $L_1 = 90$ dB SPL, for the Peltor and QuietMan headsets. The A-weighted sound level did not exceed 87 dB, to comply with Canadian Occupational Safety and Health Regulations (1991) for eight hours daily exposure. The maximum noise level was only 80 dB SPL for experiments involving the NQ100 hearing protector, because distortion (clipping) of the ANR system was observed at higher levels.

1. Stimulus control

The masked-threshold procedure described here is designed to minimize a subject's exposure to noise, by permitting exposures only to levels not greater than those experienced when the ear is fully protected by the ANR headset or hearing protector. The goal of minimizing a subject's exposure to noise was achieved by decreasing the level of the masking noise when measurements were taken with the ANR system not operating, or the headset or hearing protector not worn (unprotected ear). The external noise source was used to generate the masking noise.

The measurement procedure involves three steps (see Fig. 4). In step one, the device is worn and the ANR system is operating. The noise outside the device is kept at level L_1 , which should be the highest sound level permitted during testing. The subject is presented with a pure-tone test signal delivered through the earphone, and the masked threshold of this signal is determined by adjusting its sound pressure.

In step two, the ANR system is not operating (i.e., switched off). The important change from step one is that the test signal at the ear is now kept constant at the previously determined threshold level. The masking noise is varied to determine a new level, L_2 , which just masks the test signal. If the noise cancelling system increases the attenuation of the device, then level L_2 should be lower than the initial level used in step one (L_1). The difference in levels, $L_1 - L_2$, is a subjective measure of the active attenuation of the device.

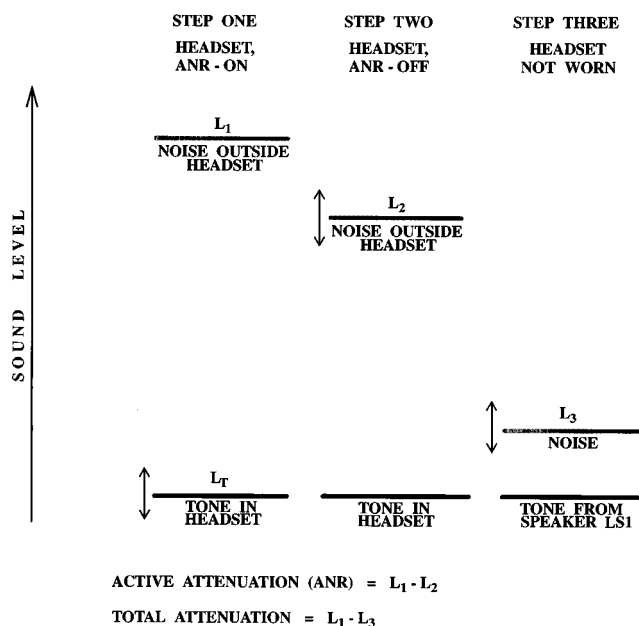


FIG. 4. Diagram of the masked-threshold procedure.

Finally, in step three, the headset or hearing protector is not worn. The test signal is presented from loudspeaker LS1, located in front of the subject. The test signal is again presented to the ear at the threshold level determined in step one of the procedure, and the level of the masking noise, L_3 , is determined that just masks the test signal. Level L_3 should be considerably lower than level L_1 , if the headset produces meaningful protection from noise. The level difference $L_1 - L_3$ is a subjective measure of the total attenuation of the device.

The level of the test signal presented in step two has to be corrected for any change in electroacoustic gain of the headset or hearing protector between the ANR-on and ANR-off conditions. Also, in step three, the gain of the loudspeaker channel LS1 has to be adjusted to obtain the same level of the test signal as the threshold level obtained in step one. The required calibrations are described in Sec. II A 3.

In the REAT method, one-third octave-band noise is recommended for the test signal, in order to obtain results integrated over a range of frequencies. For the masked-threshold method, it may be argued that using a pure-tone test signal will provide comparable listening conditions. A tone of given frequency, when presented in wideband noise, will be masked only by that part of the noise passing through the auditory filter centered at the test frequency. The auditory filter is 13% to 20% wide.¹⁶ Therefore, using a pure tone as the test signal in this masking experiment involves the integration of noise energy over a frequency range comparable to a one-third octave band (i.e., similar to that recommended for the REAT method).

2. Procedure

The experiment was carried out using an adaptive, maximum-likelihood, 2-AFC procedure, tracking 94% correct responses.^{17,18} The observation intervals were indicated by lights mounted in subject response buttons, one of which

was depressed after each pair of trials. The duration of the test tones and the interstimulus interval was 200 ms. The tones were ramped with a 10-ms sine squared function. Feedback was provided after each response, and there were 20 trials in each adaptive run.

In each step seven threshold measurements were made. The mean threshold was estimated by calculating the average of the interquartile results, that is, after excluding the lowest and the highest observed values. The mean threshold measured in step one (device worn, ANR switched on), was used to set the level of the test tone in steps two (device worn, ANR system switched off), and three (device not worn). In these steps the maximum-likelihood procedure was used to adjust the level of the masking noise.

Device attenuation was determined three times by each subject for eight test frequencies: 63, 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. During measurements, the masking noise bandwidth was reduced to two octaves below and above the test frequency. Filtering out noise bands which did not contribute to masking allowed us to limit further the subject's exposure to noise.

Stimulus presentation, timing, and response recording were controlled by an IBM-PC microcomputer equipped with a DSP card and external modules from Tucker-Davis Technologies (System II). The test tones were generated using a 50-kHz sampling rate and were low-pass filtered at 10 kHz. The pink noise was generated by a Bruel & Kjaer noise generator (model B&K 1405); its level was adjusted by attenuators controlled by the computer (module PA4 of System II).

3. Calibration

Calibration of the device's electroacoustic amplification was required to account for the differences in gain between the ANR-on and ANR-off conditions. Similarly, calibration of the loudspeaker LS1 was needed to compensate for its frequency response.

Calibration was performed during a separate session and involved physical measurement of the SPL, using a miniature microphone (Knowles BL 1785) placed in the middle of the subject's concha. The sequence of test tones used during the masked-threshold measurements was played ten times through the earphone under the two operating conditions (ANR on and ANR off), and from the loudspeaker in front of the subject. Since the position of the device on a subject's head was the most significant factor influencing the results of measurements, the subject was required to doff and don the device between measurements. This procedure increased the standard deviation of measurements but reduced bias that might result from controlled mounting of the device on a subject's head. Thus, the results are intended to represent the performance to be expected during normal use of the headset. Recordings were taken both from the left and right ears, and the data were averaged.

The differences between the SPLs recorded under the three measurement conditions were used to equalize the electrical input to the headset, hearing protector, or loudspeaker, in order to obtain the same sound pressure in the concha with the headset worn and ANR switched on. The difference in

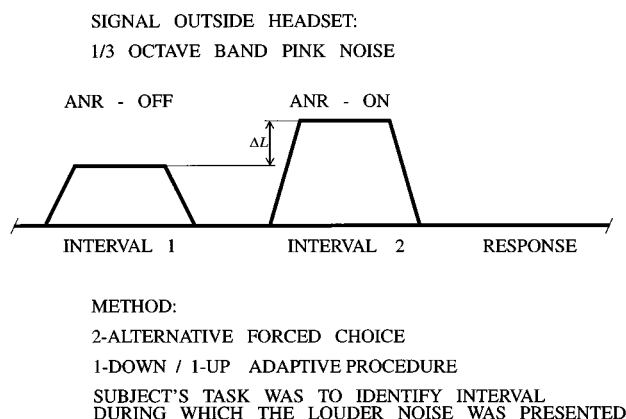


FIG. 5. Diagram of the loudness-balance procedure.

level after equalization was not greater than 3 dB for the ANR-off condition in over 95% of psychoacoustic measurements. This difference was not greater than 2 dB when loudspeaker LS1 was used. The greatest factor limiting the accuracy of the calibration procedure was the difference in gain between the ANR systems in the left and right earcups.

B. Loudness-balance method

Loudness-balance procedures offer an alternative to the REAT method for testing passive hearing protectors.^{19–21,11} In these procedures, the subject is required to compare the loudness of a test stimulus when the protector is first worn, and then not worn. The adjustment process requires the subject to don, or doff, the passive hearing protector before presentation of the next stimulus. Thus, the subject's task requires memorizing sounds while the measurement conditions are changed. In consequence, the method has been judged as having no advantages over REAT procedures, except for less demanding background noise requirements.¹¹

A loudness-balance procedure is much easier to apply to the measurement of the active attenuation of an ANR headset or hearing protector.^{9,10} The subject does not need to don, or doff, the device. Instead, the noise cancelling system is either switched on or off before the next measurement. Such a measurement protocol is less distracting to subjects and, in consequence, better accuracy may be expected.

The psychoacoustic method of adjustment has commonly been used to balance loudness.^{19–21,10} In this method, the subject changes the level of the test stimulus during the ANR-off interval until its loudness is equal to that of the stimulus presented during the ANR-on interval. When the headset or hearing protector attenuation is 40 dB or more, the sound level of a narrow-band (one-third octave) test signal may be as low as 30 dB SPL at the ear. Under these conditions, some of our inexperienced subjects reported difficulty performing the experiment.

A variant of the loudness-balance procedure was developed to avoid the problems associated with the method of adjustment. The adjustment of stimulus level was replaced by a 2-AFC, 1-down/1-up adaptive procedure. The principle of this technique is shown diagrammatically in Fig. 5.

During one of the observation intervals the ANR was switched on. During the other interval the ANR was

switched off, and the noise level outside the headset or hearing protector was changed by ΔL . The order of the ANR-on and ANR-off intervals was chosen at random. In the 2-AFC paradigm, the subject's task was to choose the observation interval during which the louder noise was presented. ΔL was changed by the adaptive procedure, which converged at 50% correct responses. This occurred when the signals reaching the ear, under the earcup, were judged of equal loudness during the ANR-on and ANR-off conditions.

Each adaptive run consisted of a sequence of 40 measurements. Feedback was provided after each response. The signal was decreased after every correct response and increased after every incorrect response. The initial step size was 5 dB, and was reduced to 2 dB after the first four reversals. The mean of signal level changes following the fourth reversal was used to estimate ΔL . The final estimate of ΔL was calculated as the mean of eight adaptive runs.

A one-third octave band of pink noise was used as the stimulus, with level corresponding to that of the 90 dB SPL broadband external noise source for the Peltor and QuietMan headsets, and 80 dB SPL for the NQ100 hearing protector. The bands of noise were centered at frequencies of 63, 125, 250, 500, 1000, and 2000 Hz for the Peltor headset. Measurements on the QuietMan and NQ100 devices also included noise bands centered at 4000 and 8000 Hz.

III. OBJECTIVE METHOD: MEASUREMENT OF INSERTION LOSS

The attenuation of passive and ANR headsets, and hearing protectors, has frequently been obtained from SPLs measured by a microphone placed in the subject's concha.^{19,22,10,7,6,23,4,8,12} Physical measurements of insertion loss, while simpler and quicker to perform than subjective measurements, do not yield results at the eardrum. A comparison between objective and subjective methods is therefore essential for active devices, which possess a secondary sound source in the proximity of the ear.

Measurements were conducted using the external noise source. The level of broadband pink noise, measured at the center head position with subject absent, was 90 dB SPL for the Peltor and QuietMan headsets, and 80 dB SPL for the NQ100 hearing protector. A block of measurements consisted of three recordings of one-third octave-band SPLs at the concha averaged over a 32-s period. Firstly, the subject wore the device and the ANR system was switched on (ANR-on condition). The ANR system was switched off during the second measurement (ANR-off condition). A third recording was made when the subject did not wear the device (unprotected ear condition). The total attenuation of the headset, or hearing protector, and that introduced by the ANR system alone were obtained by subtracting levels measured under the three conditions.

Each measurement block was repeated ten times for each subject. Subjects readjusted the position of the headset or hearing protector on the head between consecutive measurements. The left earcup was worn on the left ear and the right earcup on the right ear, corresponding to the listening condition used for the psychoacoustic measurements. Mea-

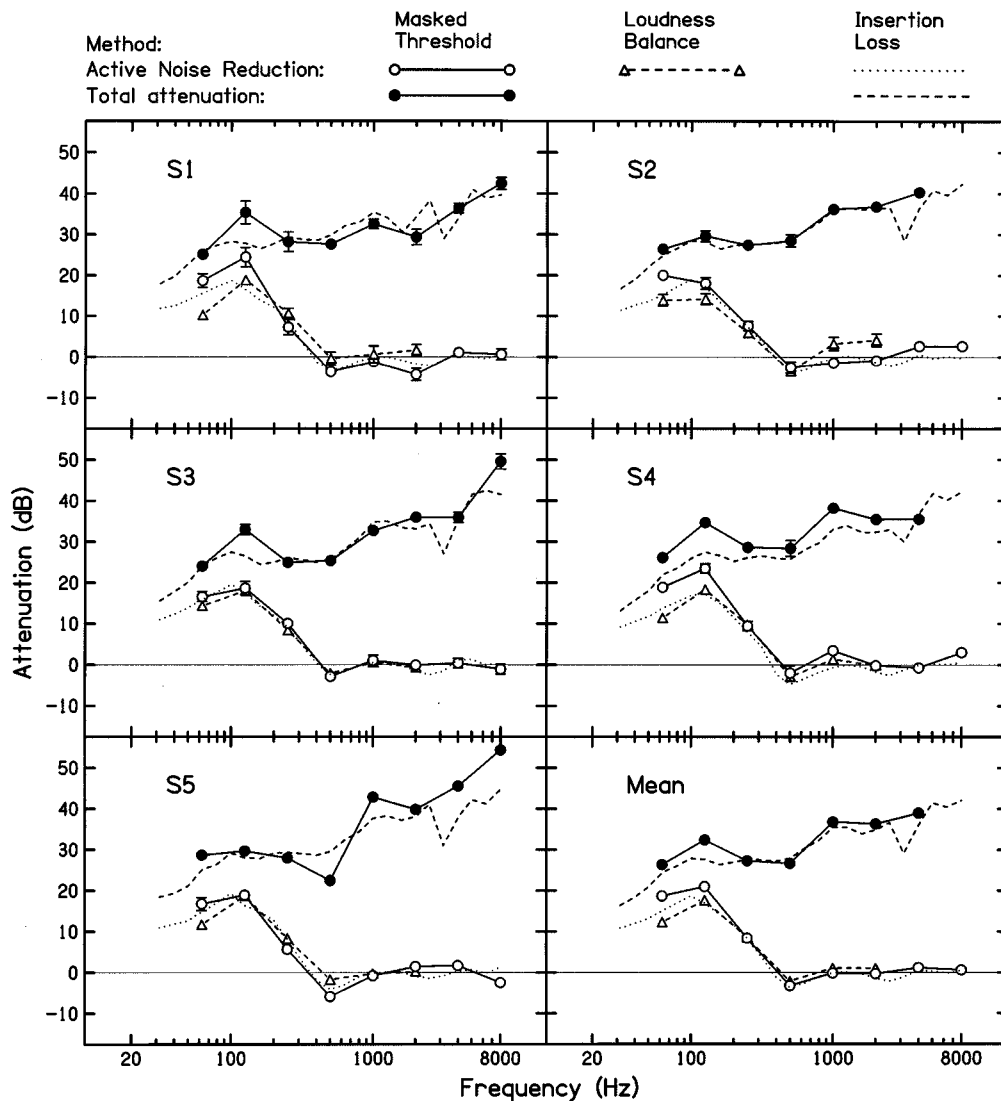


FIG. 6. Active and total (active + passive) attenuation of the Peltor headset (see text for explanation of symbols). The mean values and standard errors are shown. The average standard error is 0.46 dB for masked-threshold measurements and 0.47 dB for loudness-balance measurements.

measurements involving one device were completed during a 3-h session.

The noise was recorded by a Knowles BL 1785 ceramic microphone attached to the base of the subject's concha.²⁴ The microphone cable consisted first of thin wires (in the concha) and then of a flat cable, to avoid air leakage under the earcup seal. The signal was preamplified using a B&K 2636 measuring amplifier and processed by a B&K 2031 spectrum analyzer, where the energy equivalent SPL, L_{eq} , was computed in one-third octave bands.

IV. RESULTS AND DISCUSSION

A. Headset and hearing protector attenuation

The noise reduction obtained using subjective and objective methods is shown for the three devices in Figs. 6–8. Individual results for five subjects are shown in different panels, as well as the mean calculated from the pooled results of all subjects. The attenuation determined by the masked-threshold method is shown by continuous lines. The total ($L_1 - L_3$), and active ($L_1 - L_2$), attenuations are calcu-

lated from 70 threshold estimates and indicated by closed, and open circles, respectively. The dashed lines with open triangles show the active attenuation determined by the loudness-balance method (40 measurements per point). Finally, insertion loss measurements obtained with a microphone placed in the concha are shown as lines without symbols, the total noise reduction being indicated by a dashed line, and the ANR by a dotted line.

For the Peltor headset (Fig. 6), the total attenuation changes from approximately 20 dB at low frequencies to 40 dB at high frequencies. The total attenuation measured by the masked-threshold method is, commonly, about 1.5 dB greater than that obtained from the insertion loss measurements (see mean results). The active system of this headset is apparently designed to enhance the ANR at frequencies below 500 Hz, a maximum ANR of about 18 dB being obtained at 125 Hz. At this frequency, the masked-threshold method provided a larger estimate of attenuation than the insertion loss measurements, by about 4 dB. At 8000 Hz, it was either not possible to measure the headset's total attenu-

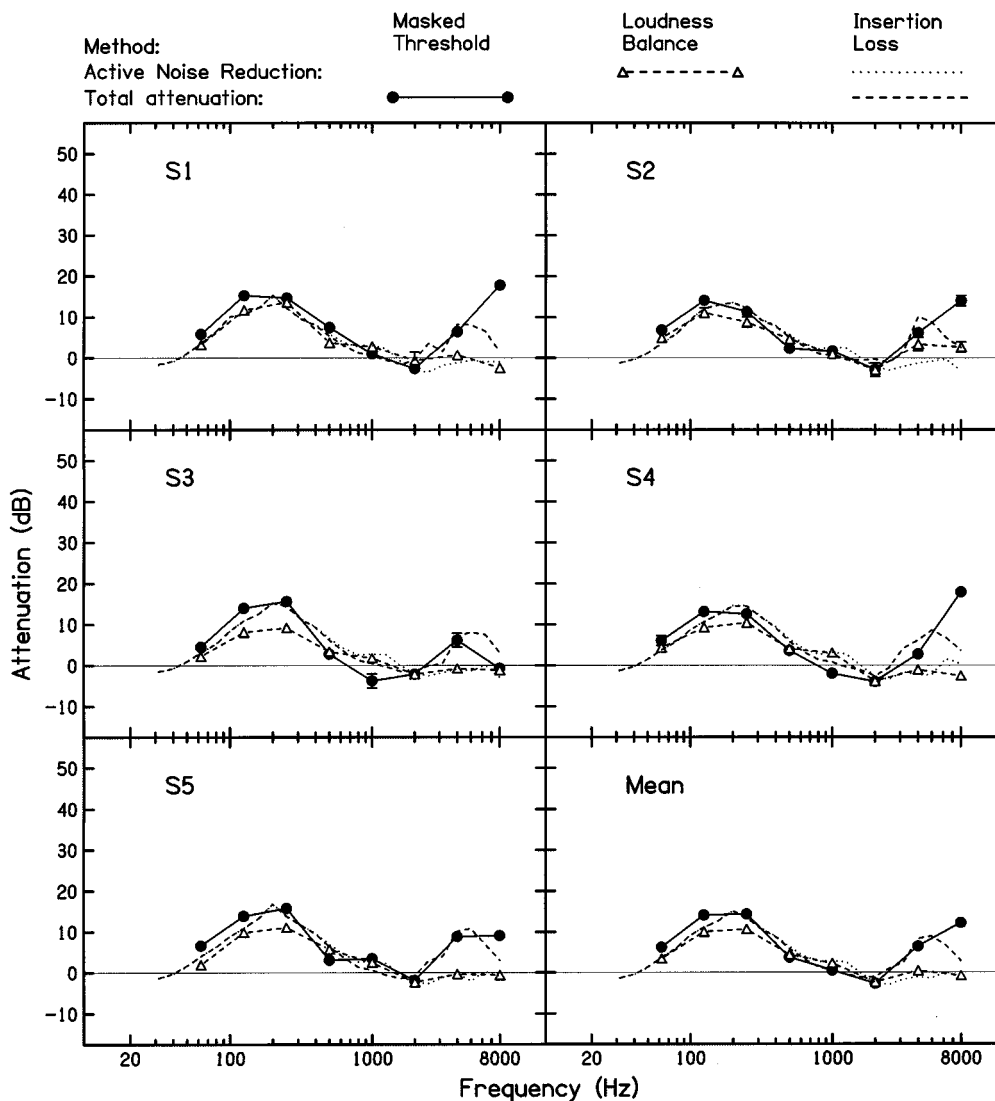


FIG. 7. Active and total (active + passive) attenuation of the NQ100 hearing protector (see text for explanation of symbols). The mean values and standard errors are shown. The average standard error is 0.35 dB for masked-threshold measurements and 0.41 dB for loudness-balance measurements.

ation by the masked-threshold method (subjects S2 and S4), or its magnitude appeared to be overestimated (subjects S3 and S5). At this frequency, the headset's total attenuation is greater than 40 dB. In effect, in step three of the procedure the test tone was presented from speaker LS1 at a very low sound level, rendering the measurement sensitive to masking by the residual noise of the multispeaker system (see Figs. 2 and 3).

The loudness-balance method provided smaller estimates of the active attenuation than the masked-threshold method. For the mean results at 63 Hz, the difference was about 6 dB, and at 125 Hz about 3.4 dB. The tendency to obtain smaller estimates of active attenuation by the loudness-balance method is consistent across all five subjects at frequencies below 250 Hz, that is at frequencies at which the attenuation is enhanced by the ANR system. We will return to this subject in a later section.

Overall, the three methods provided similar results, though some intersubject variation can be seen in the data. For subjects S1, S3, and S4, the total attenuation measured at

125 Hz by the masked-threshold method was substantially larger than that measured by the insertion loss method, by about 7 dB. A similar difference in the active attenuation at this frequency was only observed for subject S4. For subject S5, the masked-threshold method provided lower estimates of total attenuation at 500 Hz, and higher at 1000 Hz, than those obtained from insertion loss measurements. This pattern of results remained even after extensive repetitions of the masked-threshold measurements. Finally, the difference between masked-threshold and insertion loss measurements of ANR was particularly large for subject S1 at 125 Hz (8 dB), S2 at 63 Hz (4.5 dB), and S4 at 63 and 125 Hz (5 and 6.3 dB), respectively.

The attenuation of the NQ100 hearing protector is shown in Fig. 7. The design of this device did not permit a test signal to be delivered to the earphone when the device was switched off, that is, in the ANR-off condition. Thus, only total attenuation could be measured using the masked-threshold method (solid line, filled circles in Fig. 7). The active attenuation was measured by the loudness-balance

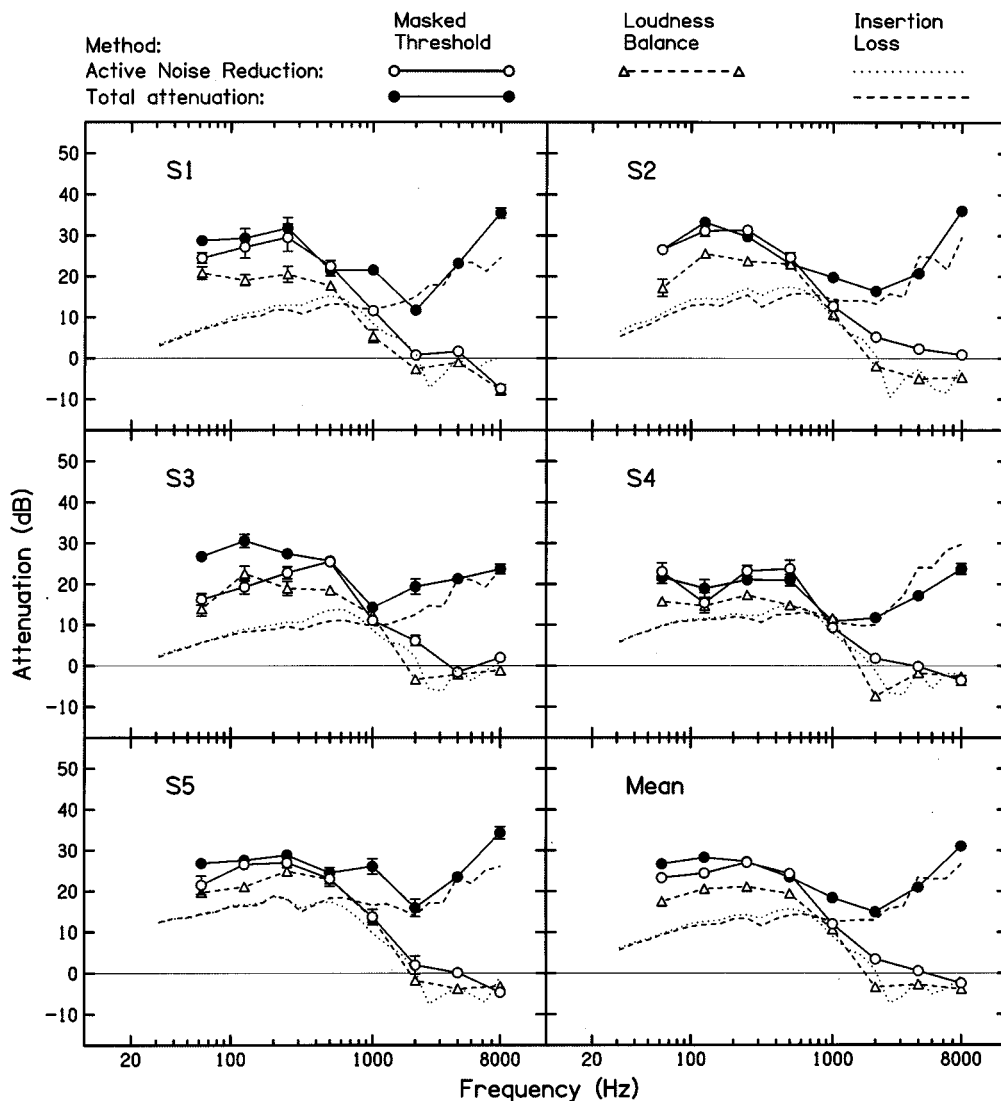


FIG. 8. Active and total (active + passive) attenuation of the QuietMan headset (see text for explanation of symbols). The mean values and standard errors are shown. The average standard error is 0.55 dB for masked-threshold measurements and 0.59 dB for loudness-balance measurements.

method (dashed line, open triangles in Fig. 7). The results show that the insertion loss reaches a maximum of about 15 dB at 200 Hz, and that passive attenuation only contributes about 10 dB to the total attenuation in the frequency range from 4000 to 8000 Hz. This performance would appear to be consistent with the lightweight, supra-aural design of the device.

The NQ100 hearing protector displayed the smallest attenuation of the three devices, and the smallest intersubject variability in the results. The masked-threshold measurements provided larger estimates of the total attenuation than the insertion loss measurements at frequencies of 63 and 125 Hz, by about 2.5 or 3 dB. Also, results obtained using the loudness-balance method were consistently smaller than those obtained using the masked-threshold method at frequencies below 500 Hz, by 3–4 dB, as observed with the Peltor headset. In addition, overestimation of the total attenuation by the masked-threshold method was again evident for almost all subjects at 8000 Hz.

An examination of the results obtained with the Quiet-

Man headset (Fig. 8) reveals that its attenuation is entirely dependent on the ANR system at frequencies below 1000 Hz. There is no significant difference between the insertion loss of the ANR system and the total (active+passive) insertion loss in this frequency range (compare dashed and dotted lines without symbols in Fig. 8). Inspection of the values obtained from masked-threshold measurements leads to a similar conclusion. At frequencies of 125, 250, and 500 Hz, the difference between the total and active attenuation is less than 3.6 dB, with the exception of subject S3 for whom this difference is about 10 dB at the lower two frequencies (see middle left panel of Fig. 8). At frequencies above 1000 Hz, the headset is essentially a passive device with total attenuation of about 15–25 dB.

Large differences between results obtained using different measuring methods were observed with the QuietMan headset. These differences occurred in the frequency range in which the headset's attenuation was generated by the ANR system. The active attenuation determined by the masked-threshold method was typically 5 dB greater than that ob-

tained from the loudness-balance method. Also, the total and active attenuation obtained by the masked-threshold method exceeded that obtained from insertion loss measurements by 15, 14, and 14 dB at 63, 125, and 250 Hz, respectively (see mean values in Fig. 8). For subjects S1, S2, and S3 this difference was as large as 20 dB at some frequencies.

Such large differences in attenuation between measuring methods cannot be attributed to random error. Note that, unlike the results obtained with the other devices, the insertion loss measurements provided the lowest estimates of active and total attenuation. Note also that the loudness-balance method generally yielded lower estimates of the active attenuation than the masked-threshold method.

B. Spatial variability of the sound field under the earcup

The QuietMan headset employs active noise cancellation operating over a wide frequency range (up to 2000 Hz), with little contribution from passive attenuation at these frequencies. In this respect the QuietMan headset is different from the other devices. Large differences between the attenuation determined by different measurement methods were commonly observed at frequencies below 1000 Hz (see Fig. 8).

The sound field under an earcup when the ANR system is operating will be formed by the superposition of the external noise transmitted to the volume enclosed by the earmuff, and the sound radiating from the earphone under the earcup.^{25,26} In this situation, rapid changes in pressure with position may occur within the earcup,^{27,7} and may be expected to occur within the ear canal. In consequence, measurements of insertion loss will depend on the position of the microphone within the earcup, and may poorly represent the pressure at the eardrum. Hence, subjective measurements may yield results that differ from insertion loss measurements obtained using a microphone in the concha.

The Quietman headset is of unusual design, involving an earphone with an annular diaphragm apparently operating as a dipole source. Also, the control system may operate in an open loop feedforward mode. It may be anticipated that in this configuration the sound-pressure variation with distance from the earphone will be large.

Additional measurements of insertion loss were conducted on two subjects (S6, S7) to test this hypothesis. A small microphone (Knowles ST497) was placed at one of three locations above the concha, or within the ear canal. The position of the microphone was controlled by adjusting the length of wires connecting it to the earphone. The wire lengths were: 5, 10, 13, 15, 18, 20, 23, or 25 mm for the QuietMan headset, and 13, 15, 18, 20, 25, or 30 mm for the Peltor headset. A wire length of 25 mm for the QuietMan headset, and 30 mm for the Peltor headset, resulted in the microphone being positioned within the ear canal, at about 15 mm from the ear canal entrance.

The distance between the microphone and earphone could not be established accurately, because the wires bent slightly when the microphone was introduced into the ear canal. It was also difficult to determine the exact position of the microphone in the ear canal, though a subjective descrip-

tion was provided by the subject. Regardless of these limitations, the measurement was sufficiently sensitive to detect changes in the microphone signal with position within the volume enclosed by the earcup.

The insertion loss obtained with the microphone at different positions close to the entrance of, and within the ear canal, is shown by the dashed lines of different length in Fig. 9. Only active attenuation was measured, and is shown for the Peltor and QuietMan headsets in the upper and lower panels, respectively. The results of loudness-balance measurements are shown for the same subjects by the continuous lines and filled circles.

At frequencies below 600 Hz, the physical measurements of insertion loss can be seen to vary by as much as 10 dB at a given frequency for the QuietMan headset, when the position of the microphone is changed. The comparable variation for the Peltor headset is only about 3 dB. At frequencies above 1500 Hz, the variation is about 2–3 dB for both headsets. Note also that, for the QuietMan headset, the subjective measurements yield estimates of active attenuation close to the largest values obtained from the physical measurements.

It is worth noting that the variability in active attenuation observed with the QuietMan headset is related to the variation in SPL recorded when the ANR system was operating. In contrast, the variation of SPL with microphone position recorded in the ANR-off condition was small, and comparable to that observed with the Peltor headset. In a separate experiment, pure tones were presented through the headset's communication channel with the ANR switched off. In these measurements the variation of SPL with location within the ear canal was not larger than 2 dB. Therefore, both experiments suggest that the variability in estimates of active attenuation is related to the sound field under the earcup when the ANR system is operating.

Note also from Fig. 9 that the active attenuation obtained with the QuietMan headset ranges from about 10 to 25 dB for subject S6, while it is only from 5 to 15 dB for subject S7. Thus, the performance of the QuietMan headset, unlike the Peltor headset, was very sensitive to the acoustical coupling between the headset and the ears of different subjects. Intersubject differences of similar magnitude can be observed in Fig. 8 for the QuietMan headset, and were larger than those obtained with the other devices. For this headset, the physical measurements of insertion loss varied by about 10 dB between subjects, the masked-threshold results by as much as 15 dB, and the loudness-balance results by about 8 dB. These intersubject variations again suggest that the same device performed differently on each person.

C. Effect of difference in ANR between left and right earcups

For all devices in this study, there was some difference in performance between the left and right ANR systems. This asymmetry will always bias binaural measurements. It will be shown that the subjective measurement methods tend either to underestimate, or overestimate, the attenuation when the ANR is different at the left and right ears. The analysis is based on the assumption that loudness depends on

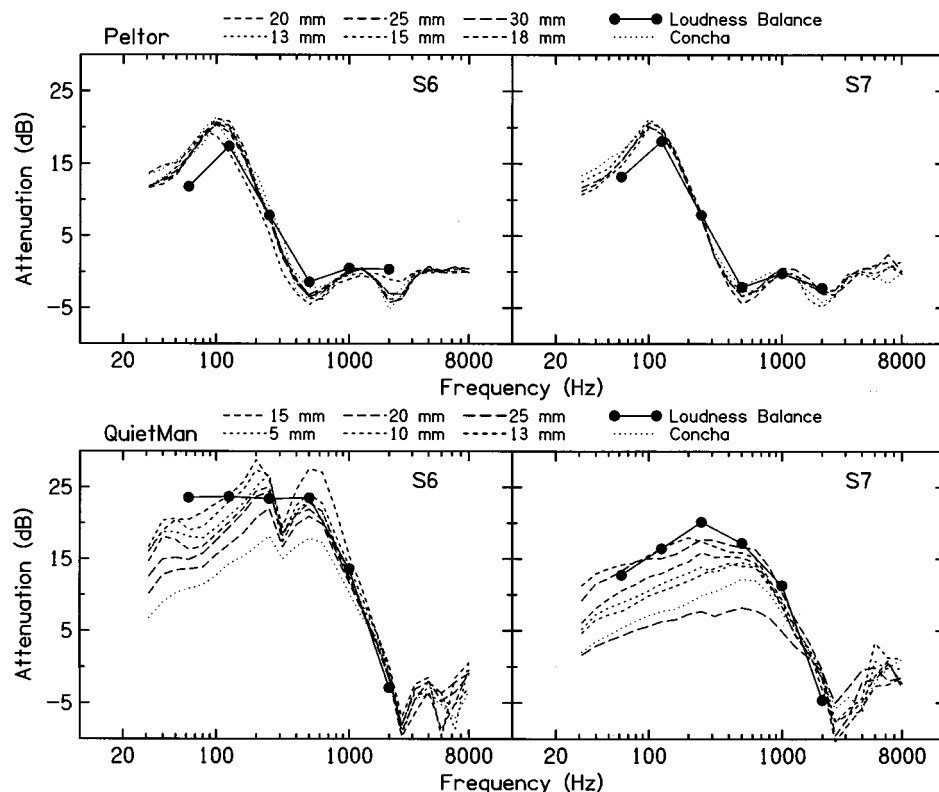


FIG. 9. Active attenuation recorded by measurements of insertion loss using different positions of the microphone under the earcup of the Peltor and QuietMan headsets. The active attenuation determined by the loudness-balance method is also shown.

the sum of the energy received by both ears. This assumption is not in disagreement with psychoacoustic data which show that, depending on SPL, a 3–8 dB increase in level is necessary under monaural presentation to equal the loudness under binaural presentation.²⁸

For a headset or hearing protector, the ear better protected will contribute little to the sensation of loudness when the difference in attenuation between earcups is large. While an asymmetry smaller than 4 dB will produce only a small change in the effective “mean” level (summed over both ears), a 6-dB asymmetry in attenuation between left and right earcups will increase the level by about 1 dB, and a 10-dB asymmetry by about 2.5 dB. If we assume that the asymmetry between left and right earcups is insignificant when the active system is switched off, the values indicate approximately how much the “mean” ANR is underestimated by the loudness-balance method.

For the masked-threshold method, any asymmetry in the performance of the ANR systems will bias the estimated attenuation towards greater values. This is because the ear that is better protected will be used to detect signals presented through the corresponding earphone. If the difference in the ANR between left and right earcups is ΔA , the masked-threshold method will give an estimate of the attenuation that is increased by about $0.5\Delta A$.

Thus, although both subjective methods will display bias resulting from the difference in attenuation between the left and right earcups, the masked-threshold method is subject to a stronger bias. In summary, the masked-threshold method will give greater active attenuation estimates than insertion

loss measurements, which are averaged over the left and right earcups, by about $0.5\Delta A$, and the loudness-balance method will give smaller active attenuation estimates, by less than $0.5\Delta A$.

An asymmetry in the performance of the ANR system also creates conditions for overestimating the total attenuation measured by the masked-threshold method. Again, the measurement affected will be that with the headset worn and the ANR switched on. Thresholds measured using speaker LS1 will not be biased because of its position directly in front of the subject (i.e., symmetrical with respect to left and right ears). Therefore, the total attenuation will again be estimated to be larger by $0.5\Delta A$ if the asymmetry in the ANR systems’ attenuation is ΔA .

Our results support the arguments presented here. For all measurements shown in Figs. 6–8, the masked-threshold method provided larger estimates of active attenuation than the loudness-balance method. For comparison, the difference between the active attenuation estimated by the two subjective methods, ΔP , as well as the difference between the insertion losses measured objectively for the left and right earcups, ΔA , are shown in Table I. The numbers are averages calculated from measurements on subjects S1–S5 at test frequencies of 63, 125, 250, and 500 Hz, that is, frequencies at which the ANR of each device was significant.

Inspection of Table I reveals that the difference between active attenuation estimated by the masked-threshold and loudness-balance methods (ΔP) increases when the difference in performance between left and right ANR systems increases (ΔA). The largest difference between results ob-

TABLE I. Average difference, ΔA , between the ANR of the left and right earcup measured physically by the insertion loss method, and the average difference, ΔP , between masked-threshold and loudness-balance methods in estimating active attenuation. Averages are computed for four test frequencies from 63 to 500 Hz.

| Headset | ΔA [dB] | ΔP [dB] |
|----------|--------------------|--------------------|
| Peltor | 1.7 | 2.1 |
| NQ100 | 2.0 | 2.5 |
| QuietMan | 4.6 | 5.1 |

tained by the masked-threshold and loudness-balance methods are seen for the QuietMan headset. For this headset the largest difference in insertion loss between left earcup and right earcups was also recorded.

Another factor that may influence the results of masked-threshold measurements is the electronic gain of the communication channel, which also drives the left and right earphones through the ANR system. Normally the gain of the communication channel changes when the ANR system is switched on, to correct the level of the incoming signal before it is modified by the ANR system. The maximum difference in gain observed was 7 dB for the Peltor headset, and about 9 dB for the QuietMan headset. These values are averages calculated from measurements on subjects S1–S5.

The calibration procedure described earlier accounted for these differences. However, the correction was calculated as the mean value for both ears. A difference in gain between left and right channels may affect the results of the masked-threshold method because there are better conditions for detecting the test signal in the channel with the greater gain. The magnitude of the corresponding error may be estimated as $(|\Delta G_{\text{ON}}| - |\Delta G_{\text{OFF}}|)/2$, where ΔG_{ON} and ΔG_{OFF} are the differences in gain between left and right channels for the ANR-on and ANR-off conditions, respectively. Thus, the error is small when the asymmetry is similar in the ANR-on and ANR-off conditions. In practice, this asymmetry in gain was about 5 dB for both ANR-on and ANR-off conditions for all headsets, so that the resulting error was less than 2 dB.

V. CONCLUSIONS

Subjective measurements of the attenuation of ANR devices involve more complicated procedures than the REAT method standardized for passive hearing protectors. The performance of ANR devices has to be measured in the presence of environmental noise and requires an additional signal source. The maximum-likelihood procedure was used to determine the thresholds in our study. Although the maximum-likelihood procedure is considered to be most efficient in estimating thresholds, other well-established psychophysical procedures for the determination of thresholds may be employed. Thus, the methods described here may be readily implemented elsewhere. Our conclusions concerning the different measurement methods studied are as follows.

(1) The masked-threshold method described here permits measurement of the active and total attenuation of an ANR hearing protector, or headset, using high noise levels. The procedure effectively minimizes a subject's exposure to

noise, and provides an opportunity to measure the attenuation under conditions closely resembling those in which the device is intended to be used.

(2) A disadvantage of the masked-threshold method is that an additional calibration is needed if the communication channel of a headset, or external loudspeaker, is used to present the test signal. Measurement of these transfer functions is a source of additional experimental error.

(3) A loudness balance method employing a 2-AFC paradigm and one-down/one-up adaptive technique is simpler and faster than the masked-threshold method described here. However, its use has to be limited to the determination of active attenuation in order to maintain its accuracy.

(4) Masked-threshold and loudness-balance methods are biased by the difference in performance of the ANR systems operating in the left and right earcups. The bias arises from the binaural presentation. The masked-threshold method will give estimates of attenuation that are larger than insertion loss measurements, averaged for both earcups, by about $0.5\Delta A$, where ΔA is the difference in attenuation between the left- and right-ear ANR systems. The loudness-balance method will give estimates of attenuation that are smaller than the average insertion loss by less than $0.5\Delta A$, if the asymmetry in insertion loss is, as found here, less than 10 dB.

(5) Measurements in this study involved a small number of subjects who tested each device sample several times. For standardization purposes it is acceptable to limit the number of measurements performed by each person on one device, and increase the number of subjects and the number of device samples. With appropriate experimental design this modification should not increase the total time required for measurements. Further research is needed to establish the influence of this strategy on the results.

(6) Performing physical measurements is faster, simpler, and less expensive than conducting psychoacoustic measurements. For these reasons, such measurements should be recommended for the assessment of headsets and hearing protectors. For ANR devices, however, the spatial variability of the sound field under the earcup may be much larger than for comparable passive devices. Therefore, the often-used technique of measuring the insertion loss on real ears with a microphone placed in the subject's concha (and the ear canal open) may give results that will not represent the attenuation experienced by the user at the eardrum. Clarification of the mechanisms responsible for the spatial variability of the sound pressure within the earcup remains a subject for future research.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Edgar A. G. Shaw for many fruitful discussions during the experiments. We also wish to thank Rene St-Denis for technical assistance, as well as John Quaroni and Sharon Sally for help in conducting the psychoacoustic tests. The work was done in collaboration with the Defence and Civil Institute of Environmental Medicine, Toronto, Ontario.

- ¹ISO 4869-1, "Acoustics—Hearing protectors—Subjective method for the measurement of sound attenuation" (International Organization for Standardization, Switzerland, 1990).
- ²ANSI S12.6, "Method for the measurement of the real-ear attenuation of hearing protectors" (American National Standards Institute, New York, 1984).
- ³CSA Z94.2, "Hearing Protectors" (Canadian Standard Association, Toronto, 1984).
- ⁴J. G. Casali and G. S. Robinson, "Narrow-band digital active noise reduction in a siren-cancelling headset: Real-ear and acoustical manikin insertion loss," *Noise Control Eng. J.* **42**, 101–115 (1994).
- ⁵D. W. Gower, Jr. and J. G. Casali, "Speech intelligibility and protective effectiveness of selected active noise reduction and conventional communication headsets," *Human Factors* **36**, 350–367 (1994).
- ⁶E. J. Walles, J. A. Lloyd, T. Ward, and S. G. Critchlow, "Non-linear hearing protector attenuation against gun-shot and industrial noise with comments on the requirements for a standard test method," *Proceedings of Inter-Noise 92* (Noise Control Foundation, Poughkeepsie, 1992), pp. 283–288.
- ⁷P. D. Wheeler and D. Smeatham, "On spatial variability in the attenuation performance of active hearing protectors," *Appl. Acoust.* **36**, 159–165 (1992).
- ⁸E. A. Goodfellow, "A prototype active noise reduction in-ear hearing protector," *Appl. Acoust.* **42**, 299–312 (1994).
- ⁹S. E. Forshaw, J. M. Rylands, and R. B. Crabtree, "Evaluating the effectiveness of active noise reduction in flight helmets," Defence and Civil Institute of Environmental Medicine, Report No. 88-RR-34 1988 (unpublished).
- ¹⁰R. B. Crabtree and J. M. Rylands, "Benefit of active noise reduction to noise exposure in high-risk environments," *Proceedings of Inter-Noise 92* (Noise Control Foundation, Poughkeepsie, 1992), pp. 295–298.
- ¹¹E. H. Berger, "Methods of measuring the attenuation of hearing protection devices," *J. Acoust. Soc. Am.* **79**, 1655–1687 (1983).
- ¹²MIL-STD-912, "Military standard: physical noise attenuation testing" (Department of Defense, USA, 1990).
- ¹³L. E. Humes, "A psychophysical evaluation of the dependence of hearing protector attenuation on noise level," *J. Acoust. Soc. Am.* **73**, 297–311 (1983).
- ¹⁴B. Scharf and S. Buus, "Audition I: Stimulus, Physiology, Thresholds," in *Handbook of Perception and Human Performance* (Wiley, New York, 1986), Vol. 1, Chap. 14.
- ¹⁵K. D. Kryter, W. D. Ward, J. D. Miller, and D. H. Eldredge, "Hazardous exposure to intermittent and steady-state noise," *J. Acoust. Soc. Am.* **39**, 451–464 (1966).
- ¹⁶B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753 (1983).
- ¹⁷C. S. Watson and D. G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Percept. Psychophys.* **33**, 113–120 (1983).
- ¹⁸D. M. Green, "Stimulus selection in adaptive psychophysical procedures," *J. Acoust. Soc. Am.* **87**, 2662–2674 (1990).
- ¹⁹J. C. Webster, "Ear defenders: Measurement methods and comparative results," *Noise Control* **1**, 34–42 (1955).
- ²⁰L. Weiner and M. L. Touger, "Variation in ear protector attenuation as measured by different methods," *J. Acoust. Soc. Am.* **32**, 245–249 (1960).
- ²¹W. Rudmose, "The case of the missing 6 dB," *J. Acoust. Soc. Am.* **71**, 650–659 (1982).
- ²²E. J. Walles, "A comparison of objective and subjective measurements of ear-muff attenuation," *Proceedings of Inter-Noise 90* (Noise Control Foundation, Poughkeepsie, 1990), pp. 1109–1112.
- ²³R. L. McKinley and C. W. Nixon, "Active noise reduction headsets," *Proceedings of the 6th International Conference on Noise as the Public Health Problem*, Nice, Italy (1993), pp. 83–86 (unpublished).
- ²⁴A. J. Brammer and J. E. Piercy, "Monitoring sound pressure within the ear: Application to noise exposure," *J. Acoust. Soc. Am.* **61**, 731–738 (1977).
- ²⁵P. A. Nelson and S. J. Elliot, *Active Control of Sound* (Academic, New York 1992).
- ²⁶P. Joseph, S. J. Elliott, and P. A. Nelson, "Near field zones of quiet," *J. Sound Vib.* **172**, 605–627 (1994).
- ²⁷P. D. Wheeler, "Voice communications in the cockpit noise environment—the role of active noise reduction," Doctoral dissertation, University of Southampton, 1986.
- ²⁸G. S. Reynolds and S. S. Stevens, "Binaural summation of loudness," *J. Acoust. Soc. Am.* **32**, 1337–1344 (1960).

Generation of zones of quiet using a virtual microphone arrangement

J. Garcia-Bonito, S. J. Elliott, and C. C. Boucher

Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, England

(Received 16 September 1996; accepted for publication 30 January 1997)

A local active noise control system is described which uses a virtual microphone arrangement. This arrangement is based on the assumption that the primary pressure at the physical and at the virtual microphone locations are similar. The implication of this assumption on the acoustic performance of a local system in a diffracted primary sound field is theoretically studied. The results show that the error at the virtual microphone position is lower when the virtual microphone arrangement is in the vicinity of a diffracting surface. A practical local active noise control system in a headrest has been built and used to measure the zone of quiet produced by a single and a dual channel system when the total pressure is canceled at one or two virtual microphone positions. It is shown that this type of arrangement is capable of projecting the zones of quiet further away from the secondary source than the position of the physical microphone. The measured zones of quiet produced by a single-channel system have been compared with the results produced with a theoretical model which includes the diffraction between two spheres, one representing the secondary source and the other the listener's head. It has been found that, in the frequency range of applicability of a local active noise control system, the two-sphere model predicts the experimental results well. The effect that the diffracting head has on the performance of an adaptive feedforward controller using a physical and a virtual microphone is also explored and the results show that this effect is small provided the initial system identification is performed in the presence of the head. © 1997 Acoustical Society of America. [S0001-4966(97)00106-9]

PACS numbers: 43.50.Ki, 43.20.Fn [GAD]

INTRODUCTION

A local active noise control system uses a secondary source to cancel the acoustic pressure at the location of an error microphone and thus generate a "zone of quiet" around this point. This type of local control system was originally described by Olson and May.¹ Ross² performed computer simulations to examine the effects of a point monopole secondary source in cancelling the acoustic pressure at another point due to a plane-wave primary sound field. He found that at low frequencies, the zone of quiet was a shell-like volume surrounding the secondary source and that as the frequency increased, the form and size of the zone of quiet degraded rapidly, being dependent on the relative orientation of the error microphone, the incident wave, and the secondary monopole. Joseph *et al.*^{3,4} investigated the zone of quiet created when the total pressure is driven to zero at a point on the axis of a piston source. By assuming a uniform pure-tone primary sound field and a feedforward control arrangement, they found that the near-field characteristics of the secondary source dictate the resulting on-axis pressure distribution. David and Elliott⁵ also performed computer simulations to estimate both the on-axis and off-axis extent of the near-field zone of quiet created by a local active control system in which the secondary source is modeled as a piston in an infinite baffle. Both uniform and diffuse models of the primary acoustic field were used. Garcia-Bonito and Elliott⁶ have recently investigated the extent of the near-field zones of quiet generated by a realistic secondary source canceling the pressure near a diffracting head in a primary diffuse sound field. They showed that, in the frequency range

of applicability of a local active control system, the distortion of the near-field zones of quiet caused by the proximity of a head is small and generally beneficial. To produce useful-sized zones, however, the error microphone may have to be positioned at an inconveniently large distance from the secondary source and may then interfere with the movement of the listener's head. Elliott and David⁷ proposed an arrangement based on the idea of a virtual microphone that can "project" the zone of quiet so that it is further away from the secondary source than the physical error microphone. This sort of arrangement avoids the interference problem between the listener's head and the error microphone, which is one of the most important limitations in any practical local active control system. In this paper we present a more complete theoretical and experimental investigation of a local active control system using such a virtual microphone arrangement.

I. A LOCAL ACTIVE NOISE CONTROL SYSTEM WITH A VIRTUAL MICROPHONE

Figure 1 illustrates the block diagram of a virtual microphone system. For single frequency excitation the complex pressure at the physical error microphone location may be written as^{7,8}

$$p_a = p_{p_a} + Z_a \cdot q_s, \quad (1)$$

where p_{p_a} is the complex primary pressure at this location, Z_a is the complex transfer impedance from the secondary source to the physical microphone location, and q_s is the

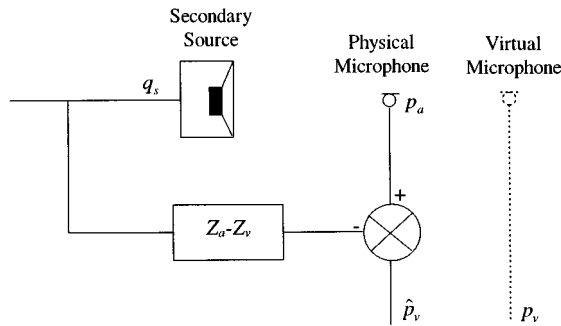


FIG. 1. A single-channel virtual microphone arrangement.

secondary source strength. Similarly, the complex pressure at a “virtual” microphone location, further away from the secondary source, can be written as

$$p_v = p_{p_v} + Z_v \cdot q_s, \quad (2)$$

where p_{p_v} is the primary field at this location and Z_v is the corresponding acoustic transfer impedance.

At low frequencies the spatial rate of change of the primary field is generally small and, therefore, we can assume that $p_{p_a} \approx p_{p_v}$. However, the acoustic transfer impedances Z_a and Z_v can be substantially different because of the intense near field of the secondary source. Under this circumstance, the pressure at the virtual microphone location can be estimated, as \hat{p}_v , from that at the physical microphone location using the following expression:⁷

$$\hat{p}_v = p_a - (Z_a - Z_v) \cdot q_s. \quad (3)$$

The estimated pressure at the virtual microphone location, \hat{p}_v , can then be canceled by the action of the secondary source using only measurements from the closer microphone, i.e., p_a . Thus knowing Z_a , Z_v , and q_s , the output from the physical microphone, p_a , can be electronically processed, as illustrated in Fig. 1, to give \hat{p}_v , which can be driven to zero by the control system, thus producing a zone of quiet around the virtual microphone location.

A. A single-channel virtual microphone arrangement

In order to study the acoustic performance of a single-channel virtual microphone arrangement in a diffuse primary acoustic field, a two-sphere model has been used in which the secondary source is modeled as an active segment in a spherical cabinet close to a rigid sphere representing the listener's head.^{6,9} In this model, the secondary pressure field is described as the superposition of those, due to an expansion of spherical harmonics about the secondary source and head positions. The coefficients of the expansions are determined by matching the velocity boundary conditions on the spheres. The pressure field can thus be calculated at all points in space including those at the physical and virtual microphone positions, Z_a and Z_v . Similarly, the primary field, including p_{p_a} and p_{p_v} , is calculated as the sum of 72 independent plane waves incident from all directions and diffracted by the two spheres.^{6,9} For each sample of primary diffuse field calculated in this way, the amplitude and phase of the secondary

source is adjusted so that the estimated pressure at the virtual microphone location, Eq. (3), is set to zero. This gives an ensemble of controlled fields which can be averaged to estimate the spatial average of the zones of quiet.

The simulated results of the average near-field zone of quiet over 20 samples of diffuse primary fields with and without a diffracting rigid sphere present are shown in Fig. 2(a) and (b), respectively. The observed lack of symmetry of the average zones of quiet with respect to the x axis is due to the limited number of diffuse acoustic field samples used in the averaging process. In these figures, the physical microphone, represented as “+,” is located at $(x,y)=(0.6L,0)$ and the virtual microphone, represented as “*”, is positioned at $(x,y)=(L,0)$, where L is the distance between the center of the secondary source and the virtual microphone position. In these figures, the secondary spherical source has a 45-degree active segment pointing in the x direction and a radius of $L/2$. The rigid sphere depicted in Fig. 2(b) has a radius of $11L/15$. These results show that the virtual microphone arrangement of Fig. 1 is effective in projecting the zones of quiet around the virtual microphone at low frequencies ($kL < 0.5$, where k is the acoustic wave number). However, at higher frequencies, the approximation $p_{p_a} \approx p_{p_v}$ is not very accurate and the extension of the zone of quiet deteriorates dramatically. We observe, for example, that for $kL=2$, which corresponds to an excitation frequency of about 1000 Hz, if the secondary source has a diameter of 0.1 m, and no head present, a 10-dB zone of quiet is not achieved around the virtual microphone. Figure 2(b) shows that the effect of a rigid sphere close to the virtual microphone improves the extension of the zone of quiet obtained, particularly at high frequencies, which is encouraging in relation to a practical application. This is caused by the fact that the pressure gradient on the surface of the rigid sphere is zero, which implies that both the primary and the secondary acoustic fields are spatially more uniform in the vicinity of the diffracting sphere.

In order to compare the effectiveness of using a virtual microphone at $(x,y)=(L,0)$ with the use of a physical microphone at the same location, the average zones of quiet have been calculated when the secondary source is adjusted to cancel the true acoustic pressure at a virtual microphone position [Eq. (2)]. Figure 3(a) and (b) show the zones of quiet created when the secondary source is adjusted to cancel the pressure measured at $(x,y)=(L,0)$ with and without the head present, respectively.⁶ Comparing Figs. 2 and 3, we observe that at low frequencies, $kL=0.2$, the 10-dB zone of quiet using a virtual microphone arrangement is very similar to the one obtained by cancelling the true pressure. At higher frequencies, $kL=1$, the 10-dB zone of quiet is substantially smaller when using the virtual microphone than with the physical microphone at the same cancellation point due to errors associated with the assumption used to derive Eq. (3). The presence of the rigid sphere turns out to be beneficial at high frequencies. We thus conclude that the difference between the zones of quiet obtained when using a virtual and a physical microphone are smaller when the rigid sphere is present [Figs. 2(b) and 3(b)] than without sphere [Figs. 2(a) and 3(a)]. This is probably due to the fact that when the rigid

sphere is present, the spatial gradient of the acoustic pressure along the x direction in the region between the secondary source and the diffracting sphere is smaller, and thus the assumption $p_{p_a} \approx p_{p_v}$ is more valid. The main conclusion that can be drawn from Figs. 2 and 3 is that a virtual microphone system is capable of projecting the zone of quiet further from the secondary source than the position of the physical microphone, making this arrangement particularly attractive for practical applications.

B. The performance of a virtual microphone arrangement in primary diffracted diffuse sound fields

The virtual microphone arrangement developed in Sec. I is based on the assumption that the primary pressure at the

physical and the virtual microphone positions are the same. In practice, however, these two pressures will be different, particularly at high frequencies. Therefore, the true attenuation achieved at the virtual microphone location is dictated by the change in the primary acoustic field between the physical and virtual microphone positions.

Figure 4 shows the arrangement considered in this work, in which a physical microphone at x is used to predict the primary complex pressure at a virtual microphone position $x + \Delta x$. We start the analysis by deriving an expression for the squared modulus of the change in the primary complex pressure between the physical and the virtual microphone positions. The change in the primary complex pressure between the positions x and $x + \Delta x$ is

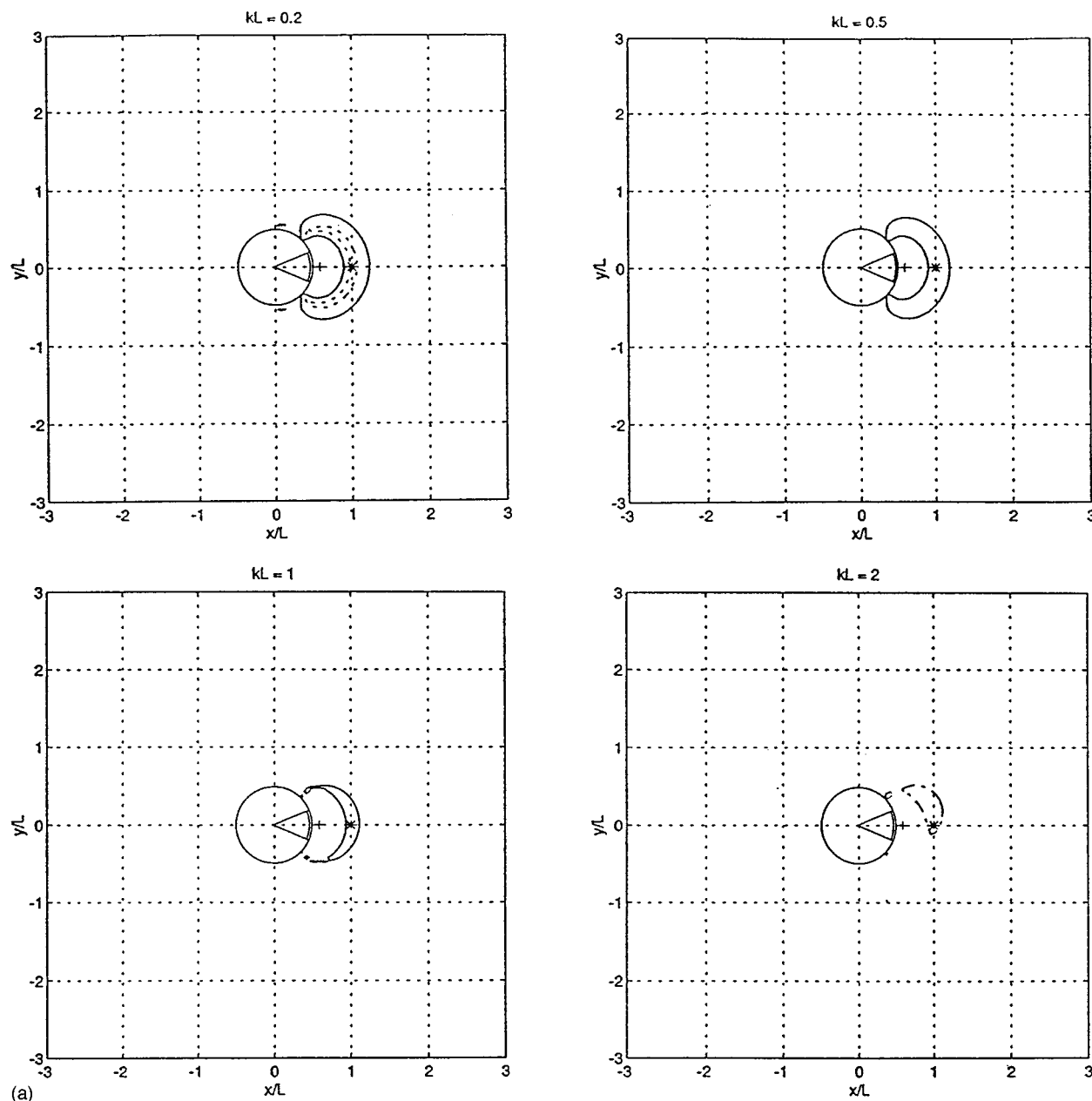


FIG. 2. (a) The calculated average zone of quiet in the x - y plane due to the superposition of a primary diffuse sound field and the field due to a secondary spherical source with a 45-deg active segment and radius $L/2$, used to cancel the pressure at a *virtual microphone* (*) located at $(x,y)=(L,0)$. The physical microphone (+) is located at $(x,y)=(0.6L,0)$. (Dash-dot line): -5 dB; (Continuous line): -10 dB; (Dotted line): -20 dB. (b) As in Fig. 2(a), but including the effect of a diffracting rigid sphere of radius $11L/15$ whose center is at $(x,y)=(2L,0)$. (Continuous line): -10 dB; (Dotted line): -20 dB.

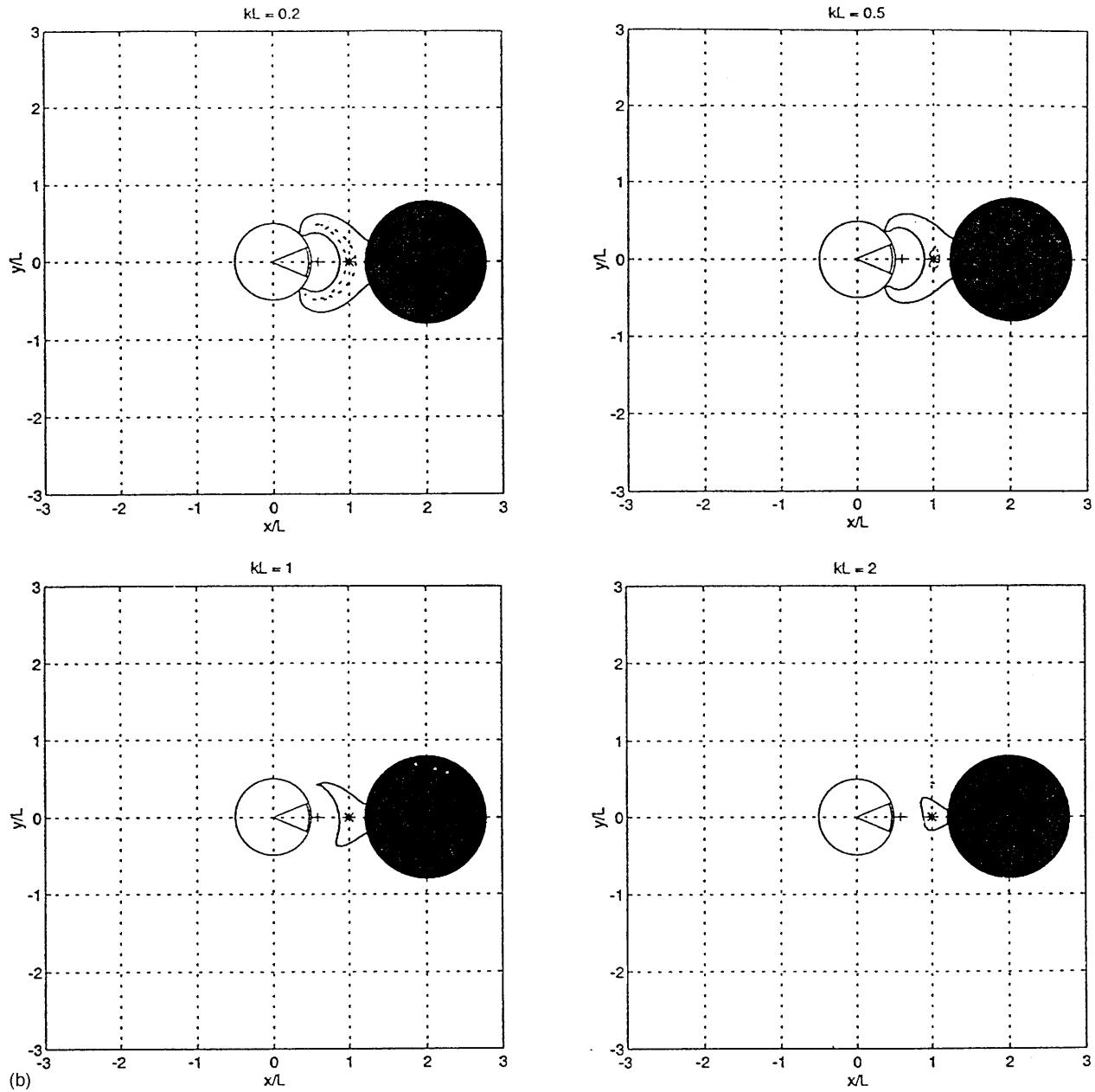


FIG. 2. (Continued.)

$$\Delta p_p = p_p(x + \Delta x) - p_p(x). \quad (4)$$

Assuming that the arrangement of Fig. 4 is immersed in a diffuse primary sound field, the primary pressure at $x + \Delta x$ can be decomposed as the sum of a component linearly related with the primary pressure at x and another component which is spatially uncorrelated,¹⁰ i.e.,

$$p_p(x + \Delta x) = p_p(x)\rho(x, \Delta x) + p_{pu}(x + \Delta x), \quad (5)$$

where $\rho(x, \Delta x)$ is defined as the spatial cross-correlation function of the primary complex pressure which is defined as

$$\rho(x, \Delta x) = \frac{\langle p_p^*(x)p_p(x + \Delta x) \rangle}{\langle |p_p(x)|^2 \rangle}, \quad (6)$$

where $\langle \rangle$ denotes spatial averaging. Substituting Eq. (5) into Eq. (4) gives

$$\Delta p_p = p_p(x)\rho(x, \Delta x) - p_p(x) + p_{pu}(x + \Delta x). \quad (7)$$

Multiplying Eq. (7) by its complex conjugate, using $\langle p_p^*(x)p_{pu}(x + \Delta x) \rangle = 0$ and taking the spatial average gives

$$\begin{aligned} \langle |\Delta p_p|^2 \rangle &= \langle |p_p(x)|^2 \rangle (\rho(x, \Delta x) - 1)^2 \\ &\quad + \langle |p_{pu}(x + \Delta x)|^2 \rangle. \end{aligned} \quad (8)$$

The spatial average of the modulus squared of the uncorrelated pressure $p_{pu}(x + \Delta x)$ is given by¹⁰

$$\begin{aligned} \langle |p_{pu}(x + \Delta x)|^2 \rangle &= \langle |p_p(x + \Delta x)|^2 \rangle - \rho^2(x, \Delta x) \\ &\quad \times \langle |p_p(x)|^2 \rangle. \end{aligned} \quad (9)$$

Substituting Eq. (9) into (8) and rearranging gives

$$\langle |\Delta p_p|^2 \rangle = \langle |p_p(x)|^2 \rangle (1 - 2\rho(x, \Delta x)) + \langle |p_p(x + \Delta x)|^2 \rangle, \quad (10)$$

which is an expression for the spatial average of the squared modulus of the change in the primary complex pressure between the physical and the virtual microphone positions in a diffuse sound field.

The spatial average value of the residual error at the virtual microphone position once Eq. (3) has been set to zero in a diffuse sound field is given by the average change in the primary acoustic field between the physical and the virtual microphone positions, i.e.,

$$\langle \Delta J_v \rangle = \langle |\Delta p_p|^2 \rangle, \quad (11)$$

where $\langle \Delta J_v \rangle$ represents the spatial average of the residual mean-squared pressure at the virtual microphone position af-

ter control. Substituting Eq. (10) in Eq. (11) and dividing both terms by $\langle |p_p(x + \Delta x)|^2 \rangle$ yields

$$\frac{\langle \Delta J_v \rangle}{\langle |p_p(x + \Delta x)|^2 \rangle} = \frac{\langle |p_p(x)|^2 \rangle}{\langle |p_p(x + \Delta x)|^2 \rangle} (1 - 2\rho(x, \Delta x)) + 1, \quad (12)$$

which is a general expression for the expected attenuation at the virtual microphone location as a function of the spatial characteristic of the primary sound field. It is interesting to note that the form of the secondary sound field plays no part in this expression, since Z_a and Z_v are assumed to be known perfectly. In general, the cross-correlation function $\rho(x, \Delta x)$ of the primary sound field is unknown and its calculation can be cumbersome. In a three-dimensional diffuse

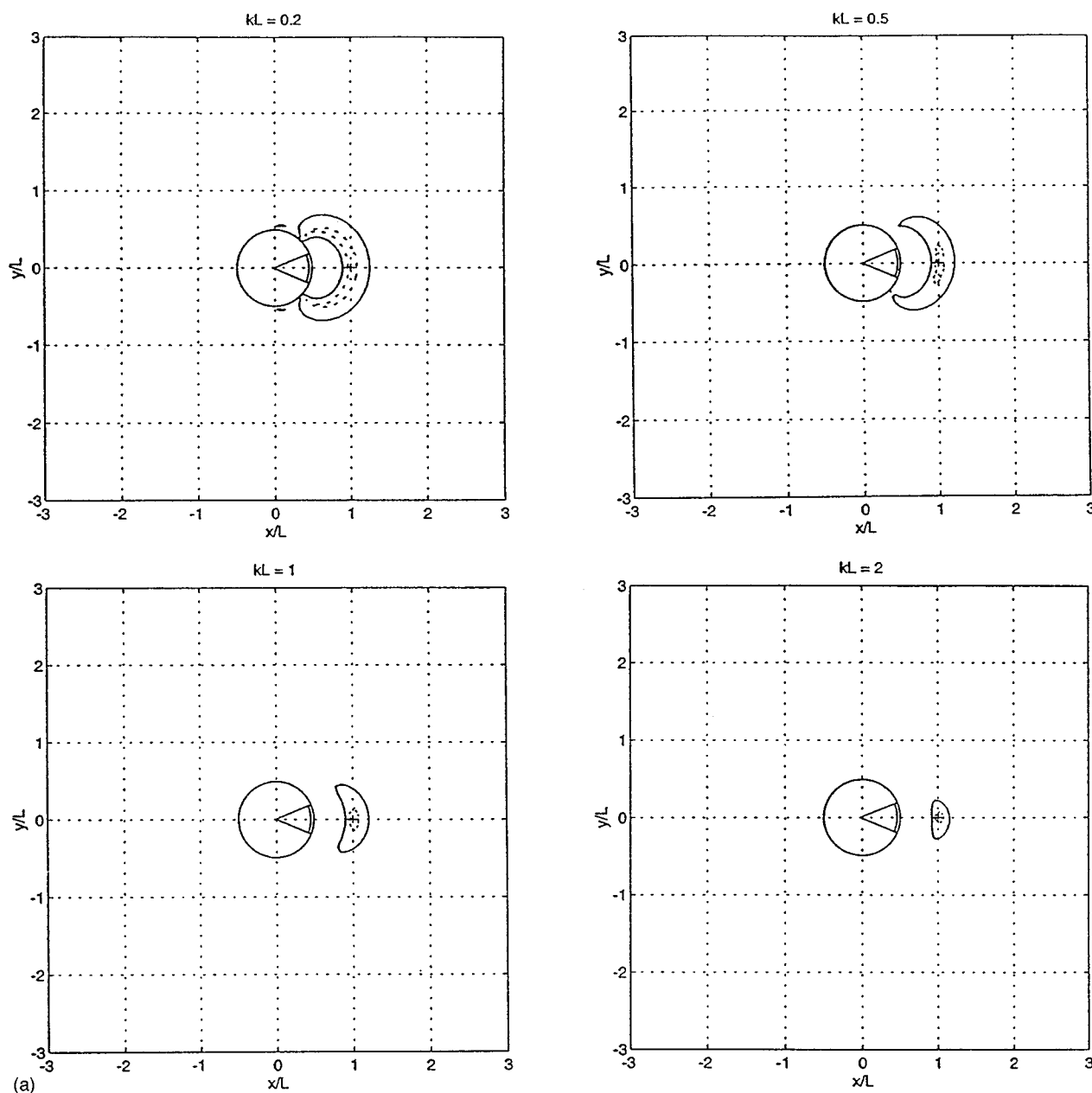


FIG. 3. (a) The calculated average zone of quiet in the x - y plane due to the superposition of a primary diffuse sound field and the field due to a secondary spherical source with a 45-deg active segment and radius $L/2$, used to cancel the pressure at a physical microphone located at $(x, y) = (L, 0)$. (Continuous line): -10 dB; (Dotted line): -20 dB. (b) As in (a) but including the effect of a diffracting rigid sphere of radius $11L/15$ whose center is at $(x, y) = (2L, 0)$. (Continuous line): -10 dB; (Dotted line): -20 dB.

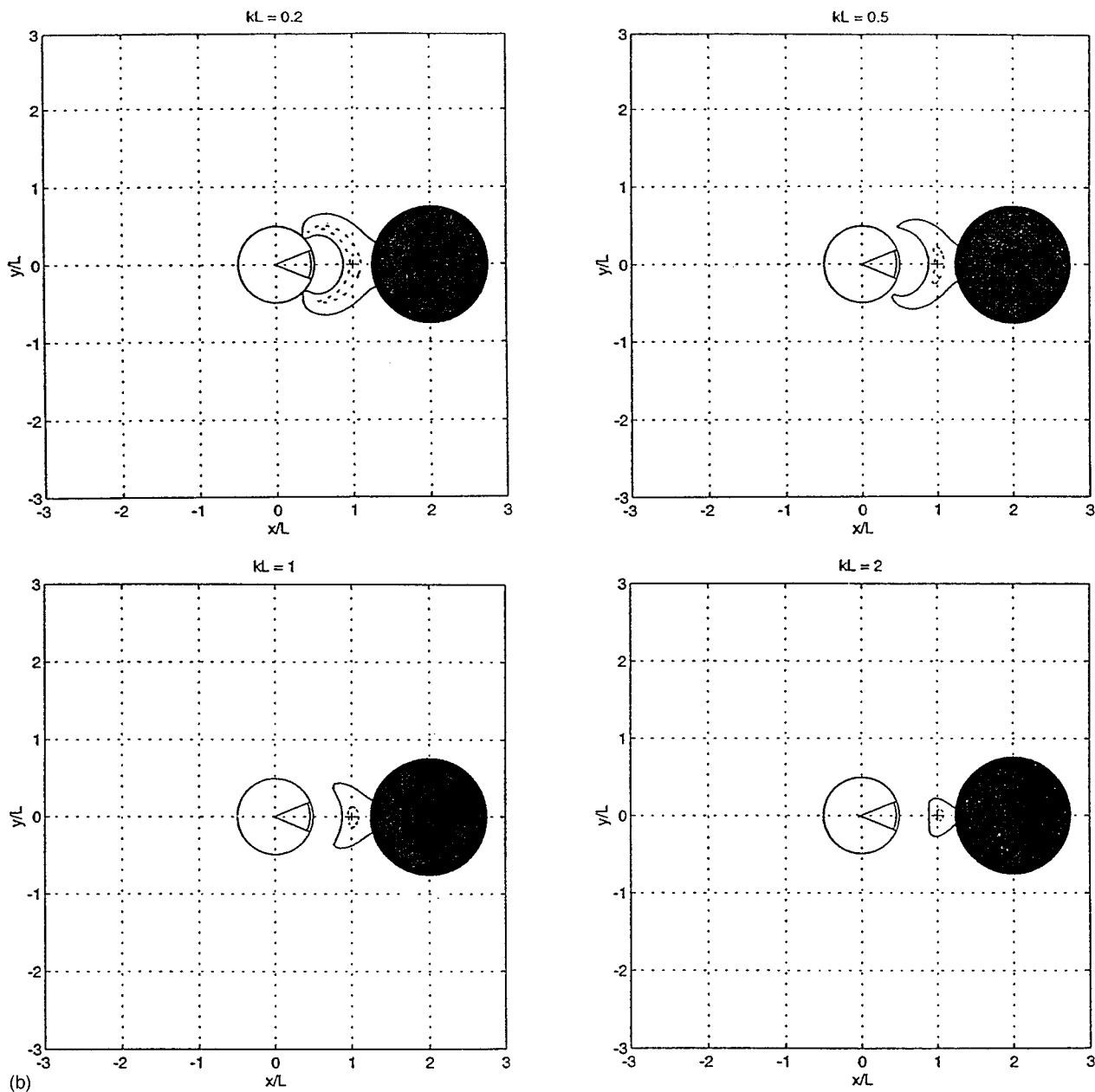


FIG. 3. (Continued.)

sound field which is not affected by diffraction, the cross-correlation function of the acoustic pressure is given by¹¹

$$\rho(x, \Delta x) = \rho(\Delta x) = \text{sinc}(k\Delta x), \quad (13)$$

where k is the wave number, and also in this case

$$\langle |p_p(x + \Delta x)|^2 \rangle = \langle |p_p(x)|^2 \rangle. \quad (14)$$

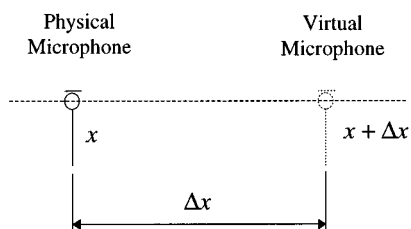


FIG. 4. A virtual microphone arrangement.

Equation (12) can then be written in this case as

$$\frac{\langle \Delta J_v \rangle}{\langle |p_p(x + \Delta x)|^2 \rangle} = 2[1 - \text{sinc}(k\Delta x)]. \quad (15)$$

If $\Delta x = 0$ in the above equation, i.e., the physical and virtual microphones are at the same position, $\langle \Delta J_v \rangle = 0$ and the cancellation is perfect, as expected. Equation (12) can however be used to predict the performance of the virtual microphone under more realistic conditions.

If the virtual microphone arrangement is located near a diffracting surface in a diffuse sound field, Eq. (13) is no longer valid and an appropriate cross-correlation function has to be derived. The performance of a virtual microphone arrangement has been calculated close to a rigid sphere, a wall, a two-wall edge, and a corner, when the physical and the virtual microphones are on the line of symmetry, i.e.,

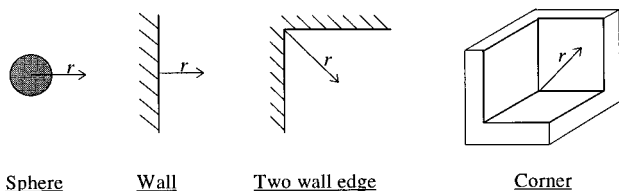


FIG. 5. The reflecting surfaces considered to study the performance of a virtual microphone arrangement near a reflecting surface in a diffuse primary sound field.

along r (see Fig. 5). This involves the use of the corresponding cross-correlation functions for a diffuse acoustic field near these types of reflecting surfaces.

The cross-correlation function of a diffuse sound field in the vicinity of a rigid sphere of radius a along the radial direction is given by¹²

$$\rho(r, \Delta r) = \frac{\sum_{m=0}^{\infty} (2m+1) \cdot C'_m(kr) \cdot C_m'^*(k(r+\Delta r))}{\sum_{m=0}^{\infty} (2m+1) \cdot [C'_m(kr)]^2}, \quad (16)$$

with

$$C'_m(kr) = (j_m^m) \cdot [\cos \delta'_{ms} \cdot j_m(kr) + \sin \delta'_{ms} \cdot n_m(kr)], \quad (17)$$

$$\delta'_{ms} = \tan^{-1} \left(-\frac{j'_m(ka)}{n'_m(ka)} \right), \quad (18)$$

where j_m denotes the spherical Bessel function of order m , n_m is the spherical Neumann function of order m and j'_m and n'_m denote their derivatives. The ratio between the average mean-squared pressure at r and $r + \Delta r$ is given by¹²

$$\frac{\langle |p_p(r)|^2 \rangle}{\langle |p_p(r+\Delta r)|^2 \rangle} = \frac{\sum_{m=0}^{\infty} (2m+1) \cdot [C'_m(kr)]^2}{\sum_{m=0}^{\infty} (2m+1) \cdot [C'_m(k(r+\Delta r))]^2}. \quad (19)$$

The corresponding correlation functions and ratios $\langle |p_p(r)|^2 \rangle / \langle |p_p(r+\Delta r)|^2 \rangle$ for a diffuse sound field near a rigid wall, a two-wall edge, and a corner along the corresponding line of symmetry represented in Fig. 5 are:¹²

For a *rigid wall*,

$$\rho(r, \Delta r) = \frac{\text{sinc}(k\Delta r) + \text{sinc}(k(2r+\Delta r))}{1 + \text{sinc}(2kr)}, \quad (20)$$

$$\frac{\langle |p_p(r)|^2 \rangle}{\langle |p_p(r+\Delta r)|^2 \rangle} = \frac{1 + \text{sinc}(2kr)}{1 + \text{sinc}(2k(r+\Delta r))}, \quad (21)$$

for a *two-wall edge*,

$$\rho(r, \Delta r) = \frac{\text{sinc}(k\Delta r) + 2 \text{sinc}(k\sqrt{(\Delta r)^2 + (2r+\Delta r)^2}/\sqrt{2}) + \text{sinc}(k(2r+\Delta r))}{1 + 2 \text{sinc}(\sqrt{2}kr) + \text{sinc}(2kr)}, \quad (22)$$

$$\frac{\langle |p_p(r)|^2 \rangle}{\langle |p_p(r+\Delta r)|^2 \rangle} = \frac{1 + 2 \text{sinc}(\sqrt{2}kr) + \text{sinc}(2kr)}{1 + 2 \text{sinc}(\sqrt{2}k(r+\Delta r)) + \text{sinc}(2k(r+\Delta r))}, \quad (23)$$

and for a *corner*,

$$\rho(r, \Delta r) = [\text{sinc}(k\Delta r) + 3 \text{sinc}(k\sqrt{2(\Delta r)^2 + (2r+\Delta r)^2}/\sqrt{3}) + 3 \text{sinc}(k\sqrt{(\Delta r)^2 + 2(2r+\Delta r)^2}/\sqrt{3}) + \text{sinc}(k(2r+\Delta r))]/[1 + 3 \text{sinc}(2kr/\sqrt{3}) + 3 \text{sinc}(2\sqrt{2/3}kr) + \text{sinc}(2kr)], \quad (24)$$

$$\frac{\langle |p_p(r)|^2 \rangle}{\langle |p_p(r+\Delta r)|^2 \rangle} = \frac{1 + 3 \text{sinc}(2kr/\sqrt{3}) + 3 \text{sinc}(2\sqrt{2/3}kr) + \text{sinc}(2kr)}{1 + 3 \text{sinc}(2k(r+\Delta r)/\sqrt{3}) + 3 \text{sinc}(2\sqrt{2/3}k(r+\Delta r)) + \text{sinc}(2k(r+\Delta r))}. \quad (25)$$

Substituting the Eqs. (20) to (25) in Eq. (12) and taking $r=a$ for the rigid *sphere* (physical microphone located at the surface) and $r=0$ for the *wall*, the *two-wall edge* and the *corner* (see Fig. 5), gives the following expressions for the expected attenuation at the virtual microphone location in a diffracted diffuse sound field: For a *sphere*,

$$\frac{\langle \Delta J_v \rangle}{\langle |p_p(a+\Delta r)|^2 \rangle} = \frac{\sum_{m=0}^{\infty} (2m+1) (C'_m(ka))^2 - 2 \sum_{m=0}^{\infty} (2m+1) C'_m(ka) C_m'^*(k(a+\Delta r))}{\sum_{m=0}^{\infty} (2m+1) (C'_m(k(a+\Delta r)))^2} + 1 \quad (26)$$

for a *wall*,

$$\frac{\langle \Delta J_v \rangle}{\langle |p_p(\Delta r)|^2 \rangle} = \frac{2(1 - 2 \text{sinc}(k\Delta r))}{1 + \text{sinc}(2k\Delta r)} + 1, \quad (27)$$

for a *two-wall edge*

$$\frac{\langle \Delta J_v \rangle}{\langle |p_p(\Delta r)|^2 \rangle} = \frac{4(1 - 2 \text{sinc}(k\Delta r))}{1 + 2 \text{sinc}(\sqrt{2}k\Delta r) + \text{sinc}(2k\Delta r)} + 1, \quad (28)$$

and for a *corner*,

$$\frac{\langle \Delta J_v \rangle}{\langle |p_p(\Delta r)|^2 \rangle} = \frac{8(1 - 2 \operatorname{sinc}(k\Delta r))}{1 + 3 \operatorname{sinc}(2k\Delta r/\sqrt{3}) + 3 \operatorname{sinc}(2\sqrt{\frac{2}{3}}k\Delta r) + \operatorname{sinc}(2k\Delta r)} + 1. \quad (29)$$

Figure 6 shows the average attenuation at the virtual microphone in a single channel local active control system in a diffuse sound field with the *physical* microphone located on the surface of a rigid sphere with $ka = 1.5$, (a) [Eq. (26)], on a rigid wall, (b) [Eq. (27)], on an edge, (c) [Eq. (28)] and on a corner, (d) [Eq. (29)]. The horizontal axis represents the distance between the physical and the virtual microphone normalized to the wavelength. These curves are compared with the average performance that it would be obtained if the arrangement was located in a nondiffracted diffuse sound field, (e) [Eq. (15)]. These results show that the diffraction effects near the virtual microphone arrangement considerably improve the acoustic performance of the local control system. In practice, a local active control system with a virtual microphone arrangement implemented in a headrest, for example, will be affected by the diffraction of the headrest and the listener's head. The results in Fig. 6 show that as the extent of the reflecting surfaces increase, compared with the acoustic wavelength, the attenuation achieved at the virtual microphone is expected to increase. This is due to the fact that the primary diffuse field becomes spatially more uniform as the size of the reflecting surface increases.¹² As the separation between the physical and the virtual microphone increases with respect to the wavelength, however, the assumption that the primary acoustic pressure at the virtual microphone location is the same as that at the physical microphone becomes increasingly inaccurate, which causes a reduction in the attenuation achieved at the virtual microphone position. It is interesting to note that for a separation greater than about 0.3λ , the controlled pressure at the virtual

microphone position is greater than the primary pressure for all the cases considered and that this increase can be up to 20 dB for the case of a corner and $\Delta r = 0.5\lambda$. The maximum distance between the physical and virtual microphone at which 10 dB of attenuation can still be achieved at the virtual microphone location has been increased by the presence of the diffraction from about $\Delta r = 0.1\lambda$ in free space to about $\Delta r = 0.2\lambda$.

II. A PRACTICAL ACTIVE HEADREST

The size of a zone of quiet created by a local active control system is greatest when the active surface of the secondary source is large and when the cancellation point is remote from the secondary source.^{3,5} This situation presents severe limitations when a practical implementation of a control system is considered. In the practical case, relatively small loudspeakers are required due to restrictions on the available space. Additionally, in order to avoid physical interference between the listener's head and the error microphones, these microphones have to be located away from the person's head and relatively close to the secondary sources. These geometric constraints limit the size of the zones of quiet and the attenuation at the listener's ears achievable with a conventional local active noise control system. In Sec. II it was demonstrated that the virtual microphone configuration is capable of projecting the zones of quiet further away from the secondary source than the position of the error sensor. This allows us to create a zone of quiet near a listener's ears without the need for locating the error microphones at an inconvenient position and, at the same time, reap the benefit from having the cancellation points relatively distant from the secondary sources, which is known to produce bigger zones of quiet.

In this section we investigate the zones of quiet that can be obtained with a realistic local active noise control system using an adaptive feedforward controller installed in the headrest of a passenger seat. A single-channel and a dual-channel control system using physical and virtual microphones have been investigated experimentally. A two-sphere model has also been used to theoretically predict the dimension of the zones of quiet for the single-channel case.⁶

The effect of a diffracting head on the performance of a local active noise control system using a feedforward controller with a virtual microphone is then considered using the two-sphere model presented above.

A. Design of a local active noise control system in a headrest and experimental setup

In order to explore the performance that can be obtained with a practical local active noise control system implemented on a passenger's seat, the headrest assembly shown in Fig. 7 was built. This replaced the normal headrest in an individual seat designed for train passengers which was

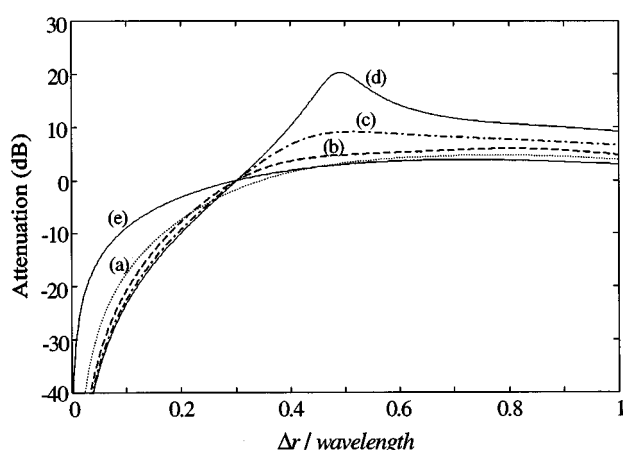


FIG. 6. The average attenuation at the virtual microphone position for a single-channel local active noise control system in a primary diffuse sound field when the physical microphone is located on the surface of a rigid sphere with $ka = 1.5$ (a), a wall (b), a two-wall edge (c), and a corner (d). The attenuation for the case of a nondiffracted diffuse field is shown in curve (e). Δr is measured along the radial direction for the case of the sphere; in the normal direction, for the wall, and along the line of symmetry for the case of the edge and the corner (see Fig. 5).

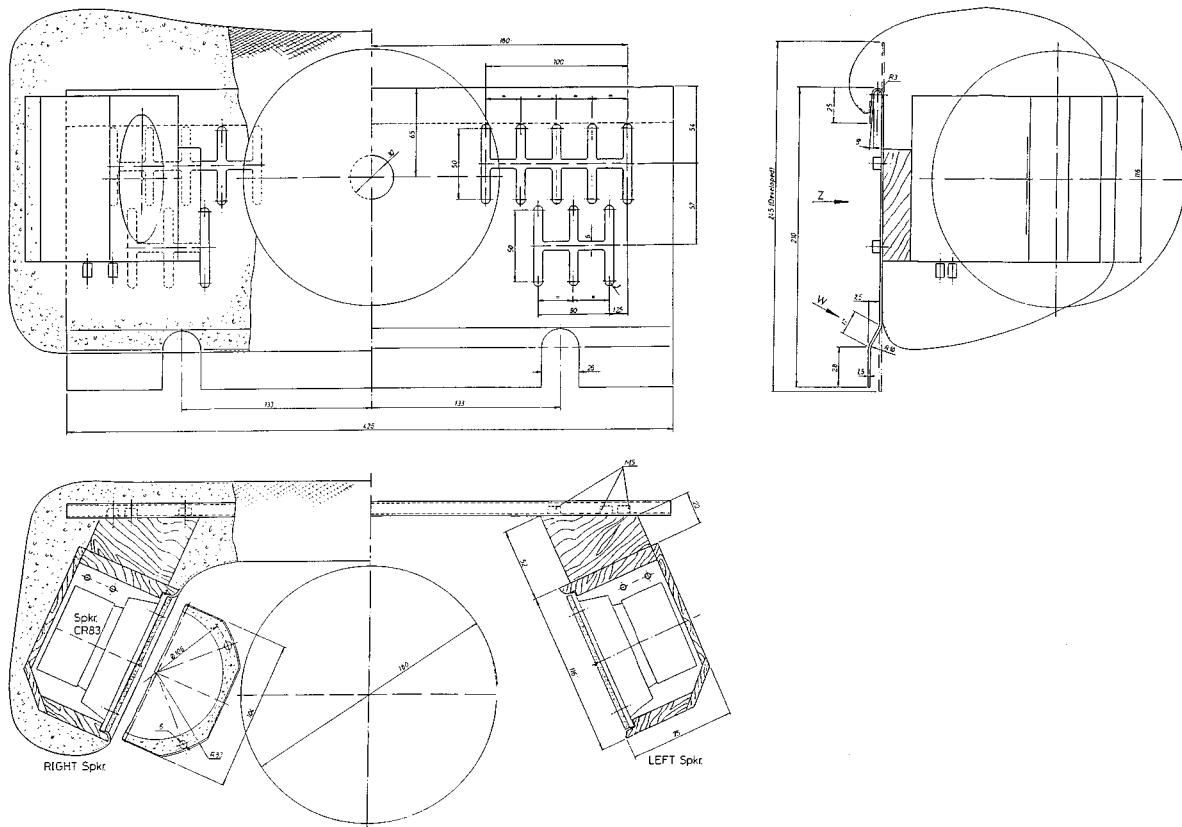


FIG. 7. The dual channel local active noise control system in a headrest used for the measurements.

available for laboratory use. In this figure, two loudspeakers with a diaphragm diameter equal to 100 mm, were bolted to an aluminum panel through a set of distributed slots so that it could be adjusted vertically and horizontally. In the middle of the headrest a sphere is shown to illustrate the average size of a human head and to estimate the approximate magnitude of the possible head movements allowed by this particular design. The padding shown in the drawing was not fitted in the experimental arrangement to give greater freedom in adjusting the position of the loudspeakers.

Figure 8(a) shows the headrest mounted on a passenger seat and a manikin with an aluminum head and a wooden torso which was designed specifically for the experiments reported here. The manikin was built so that its overall dimensions were within 6% of the average human adult head and torso.^{13,14} In order to monitor the acoustic pressure heard by this hypothetical listener, an electret pressure microphone was located at the entrance of both ear canals of the manikin head. It is known that measurement of the acoustic pressure at the entrance of the ear canal introduces an error of less than 1.5 dB compared with that at the eardrum for frequencies of less than 1 kHz.¹⁵ Figure 8(a) also shows two adjustable brackets used to position the two error microphones in the desired location. The two adjustable microphones were only allowed to move on a horizontal plane containing the centers of the loudspeakers and the centers of the ear canals of the manikin. The manikin rested on a horizontal surface at the level of the armrests which allowed the manikin to be moved horizontally to various positions. Figure 8(b) is a sketch of the active headrest of Fig. 8(a) illustrating the rela-

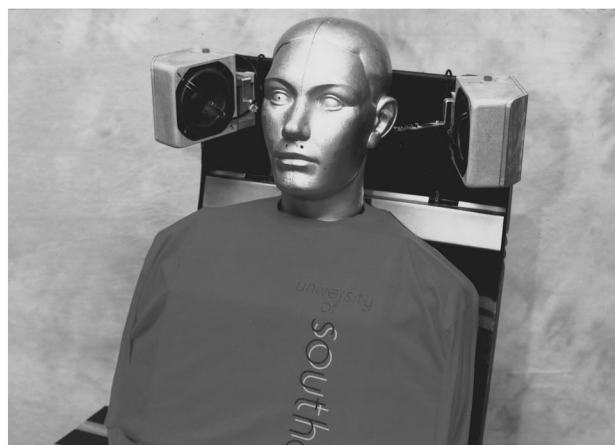
tive position of the listener's head and the error microphones as well as the reference system used to define the position of the manikin.

B. Measurements on a two-channel system

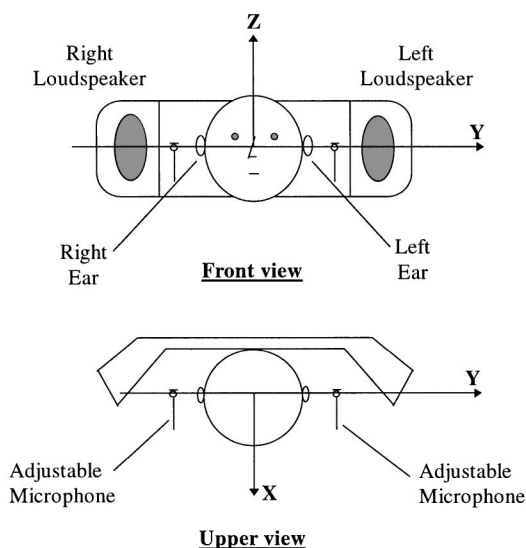
1. A two-channel system using physical microphones

In order to carry out measurements of the zones of quiet achievable by the practical local active noise control system considered here, the assembly shown in Fig. 8(a) was located in a listening chamber of dimensions $5 \times 2.2 \times 2.5$ m which had reasonably anechoic characteristics above 200 Hz. The seat was located facing in the direction of the largest dimension of the room and a primary loudspeaker was positioned 2 m behind it, radiating noise towards the back of the seat. In this way, the primary acoustic field will approximate a progressive plane wave moving in the positive direction of the x axis [see Fig. 8(b)].

For each frequency of interest, the room was excited using the primary loudspeaker to create the primary acoustic field. The phase and amplitude of the two secondary loudspeakers were then electrically adjusted to cancel the primary acoustic pressure measured by the two error microphones. The control was carried out in real time using a two-channel ANC program developed and implemented in a DSP. The signal driving the primary loudspeaker was generated by the control program so that the frequency of the signals feeding the secondary and primary loudspeakers were synchronized. The control program is based on the LMS algorithm and, thus requires the identification of the frequency responses



(a)



(b)

FIG. 8. (a) The headrest mounted on a passenger's seat and the manikin used in the measurements of the realistic zones of quiet. (b) Sketch of the active headrest of (a), illustrating the relative position of the secondary sources, listener's head, and error microphones. The origin of the reference system corresponds to the center of the head (which is assumed to be spherical) when it is laterally centered and in contact with the headrest.

between the secondary loudspeakers and the error microphones. This identification stage, which must be done before applying control, can be performed at the fundamental frequency of excitation, which is selected by the user, and at the first seven harmonics at the same time. In all the measurements reported here, the identification of the acoustic path was performed with the head laterally centered with respect to the headrest and in contact with it. The frequency responses measured in this way were used by the control program to minimize the error signal for every other position of the head.

Figure 9 shows the attenuation measured at the right and left hand ears when the two secondary loudspeakers were adjusted to cancel the primary acoustic field at the two error microphones located close to the secondary loudspeaker but 9.5 cm from the manikin's ears on a line through the ears, for different positions of the manikin along the x axis. As expected, due to the symmetry of the physical arrangement, the

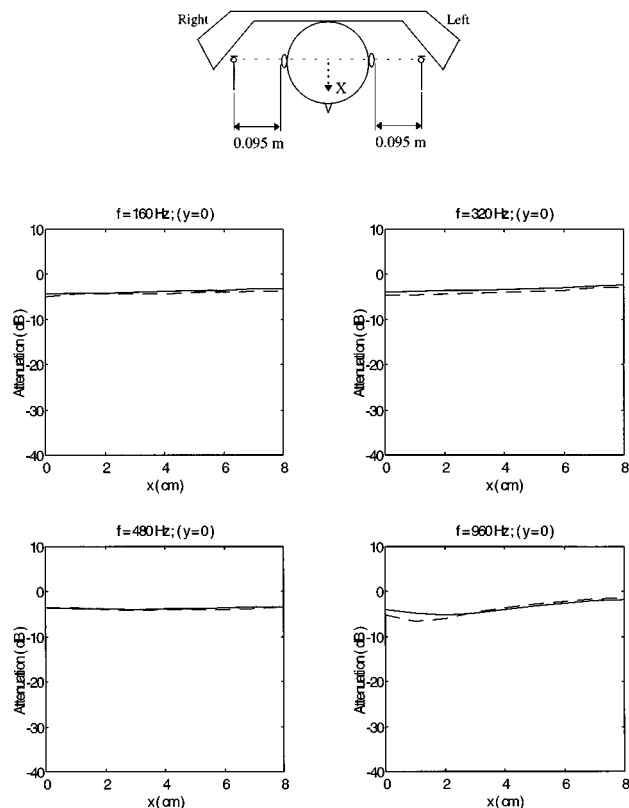


FIG. 9. The measured controlled acoustic field at the right- (continuous) and left-hand ears (dashed line) when the two secondary loudspeakers cancel the primary acoustic field at the two *close* error microphones, for different positions of the head and torso along the x axis.

attenuation measured at both ears is almost the same and is very low due to the considerable distance between the microphones and the ears. Figure 10 is similar to Fig. 9 but for positions of the manikin along the y axis. These results show that the microphone arrangement used in Figs. 9 and 10 gives poor attenuation at the listener's ears, particularly when the head is near the center of the headrest.

Figure 11 shows the attenuation measured at the right- and left-hand ears when the two secondary loudspeakers were adjusted to cancel the primary acoustic field at two error microphones located at 2 cm from the ears, for different positions of the manikin along the x axis. These results show that in order to obtain a good attenuation at the listener's ears the cancellation point must be very close to them. The difference in the measured attenuation at the two ears at a frequency of 160 Hz may be due to the influence of a room mode in the listening room. At high frequencies the room is almost anechoic and the attenuation measured at both ears is very similar. It was not possible to move the manikin in the y direction with this arrangement due to the presence of the error microphones.

2. Two-channel system using virtual microphones

The experimental procedure followed in the case of a local active noise control system with virtual microphones is similar to that described in the previous section for the physical microphones arrangement. The only difference is that the control program requires an identification stage during which

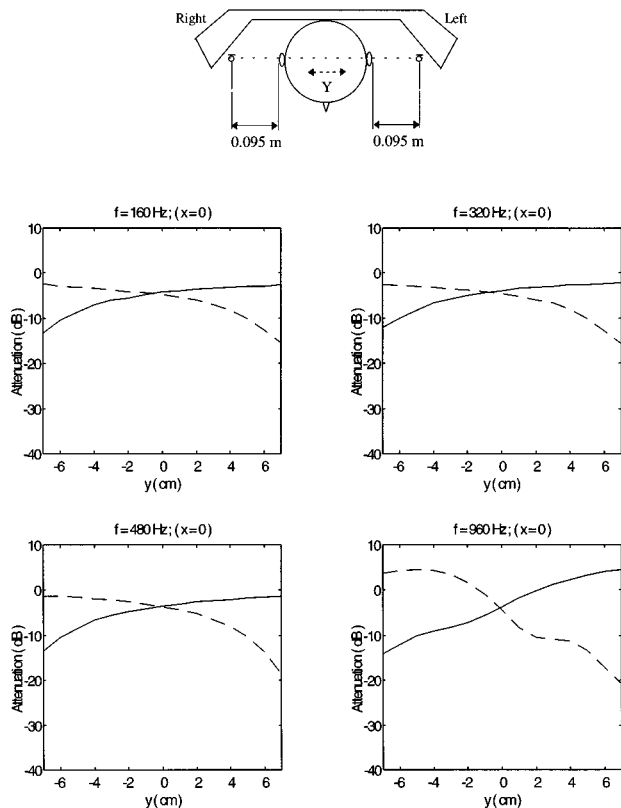


FIG. 10. The measured controlled acoustic field at the right- (continuous) and left-hand ears (dashed line) when the two secondary loudspeakers cancel the primary acoustic field at the two *close* error microphones, for different positions of the head and torso along the y axis.

the frequency responses between the secondary loudspeakers and the microphones located at the virtual sensor positions were measured. In order to implement Eq. (3) in the control system, these frequency responses were then stored and subtracted from the frequency responses between the same loudspeakers and the physical microphones positioned at the physical sensor locations. These measurements were used to implement the two-channel version of the virtual microphone arrangement presented in Sec. I. It is important to mention that these responses were measured prior to control with the manikin in the central position and the same responses were then used by the control program for all manikin positions.

Figures 12 and 13 show the measured attenuation at the manikin's ears using a two-channel virtual microphone arrangement for different positions of the manikin in the x and y directions, respectively. The relative position of the physical and the virtual microphones is shown in the same figures. This implementation is an example of a practical active local active noise control system in which the physical error microphones are in a noninterfering position and the zones of quiet are projected away from the secondary sources and close to the listener's ears. It is important to note that Fig. 13 could not be obtained with the arrangement of Fig. 11 because the two physical microphones restrict the head movement in the y -direction. It is also interesting to note that in spite of the fact that the virtual microphone location is 2 cm from the ears, the deepest minimum in Fig. 13 is 1 cm away from the ears. The reason for this displacement of the zone of

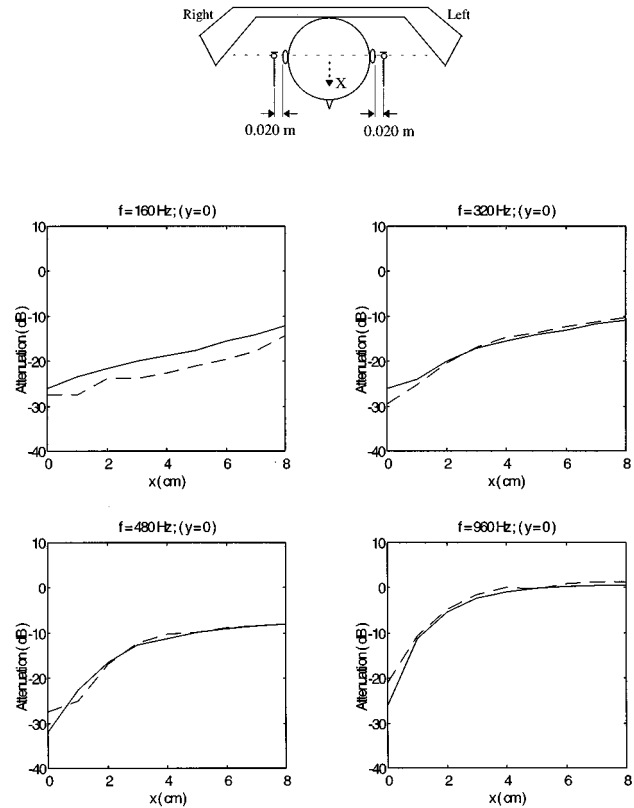


FIG. 11. The measured controlled acoustic field at the right- (continuous) and left-hand ears (dashed line) when the two secondary loudspeakers cancel the primary acoustic field at the two *remote* error microphones, for different positions of the head and torso along the x axis.

quiet is that as the head moves towards the virtual microphone position, the acoustic transfer impedance between the secondary source and this position increases. This has the effect of projecting the zone of quiet away from the secondary source which agrees with the experimental results shown in Fig. 13.

The performance of the virtual microphone arrangement of Fig. 12 should be compared with Fig. 11 for which the physical microphones were located close to the manikin's ears and therefore is the arrangement which the virtual microphone system is trying to mimic. At low frequencies, the primary acoustic field at the physical and virtual microphone locations are very similar and the control system can calculate an accurate estimate of the total pressure at the virtual microphone position. This means that the controller can produce almost perfect cancellation at the virtual microphone position. At higher frequencies, however, the difference in the primary acoustic field between the physical and the virtual microphone positions limits the attenuation achievable at the virtual microphone position.

The measured results indicate that the virtual microphone arrangement produces useful zones of quiet around the virtual cancellation point up to frequencies of about 500 Hz. This frequency is well above the frequency at which most practical global control systems cease to be effective, which suggests that the use of virtual microphones might have important practical applications for local active noise control at high frequencies. We also notice that the use of a

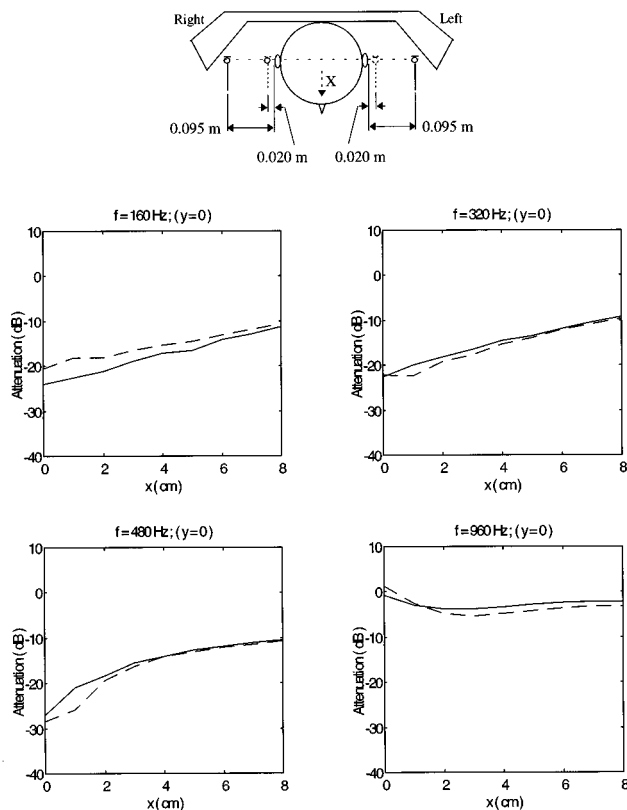


FIG. 12. The measured controlled field at the right- (continuous) and left hand-(dashed line) ears when the two secondary loudspeakers are adjusted to cancel the primary acoustic field at the *two virtual microphone locations*, for different positions of the head and torso along the *x* axis. The primary acoustic field is a plane wave propagating in the positive *x* direction.

virtual microphone allows lateral displacements of the head in the direction towards the secondary sources without significantly distorting the symmetry of the zone of quiet with respect to the virtual cancellation point and that the 10 dB zone of quiet extends to approximately 8 cm forward and 10 cm from side to side up to a frequency of 500 Hz. This result is a strong motivation to use this virtual microphone arrangement in practical applications.

C. Theoretical predictions on a single-channel system

The physical system shown in Fig. 8 represents a 2×2 channel system seeking to cancel the acoustic pressure at two error microphones. Because the pressure generated by any one secondary loudspeaker at the closest error microphone is considerably greater than that at the error microphone located further away from the loudspeaker, the two secondary sources behave almost independently when control is applied. This means that the source strength associated to each loudspeaker to achieve pressure cancellation at the two error sensors will be mainly dictated by the acoustic transfer impedances between each secondary source and the closest error sensor. This assumption allows us to model the right or left half of the arrangement shown in Fig. 8 as a single-channel system. On the other hand, since we are mainly concerned with the generation of zones of quiet in a relatively low-frequency range, say up to 1 kHz, the actual geometric

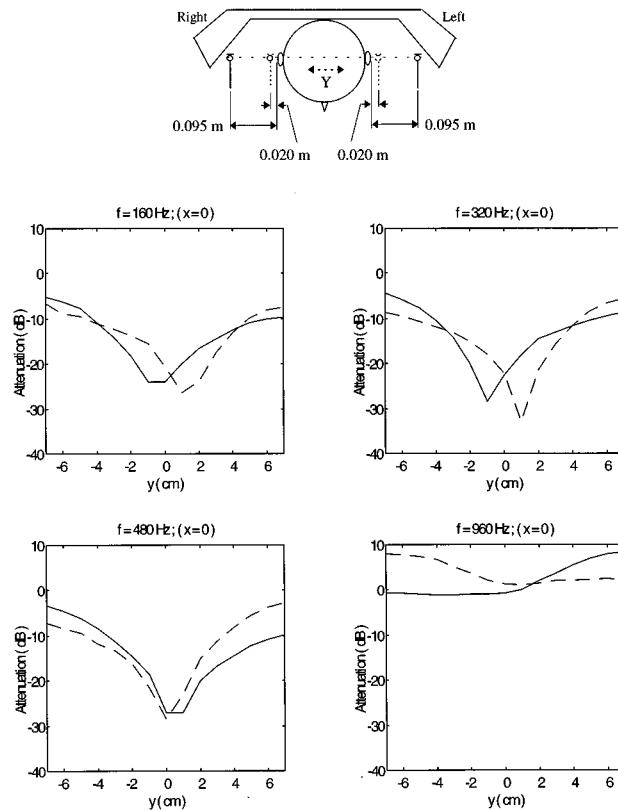


FIG. 13. The measured controlled field at the right (continuous) and left hand (dashed line) ears when the two secondary loudspeakers are adjusted to cancel the primary acoustic field at the *two virtual microphone locations*, for different positions of the head and torso along the *y* axis. The primary acoustic field is a plane wave propagating in the positive *x* direction.

details of the headrest shown in Fig. 8 may be neglected and we can assume that, for a particular error microphone location, the generation of the near-field zone of quiet is mainly dictated by the relative size and proximity of the secondary source and the diffracting head. Under these assumptions, a two-sphere model in which the secondary loudspeaker is modeled as an active segment in a spherical cabinet and the listener's head is modeled as a rigid sphere, can be used to try to predict the zone of quiet created by the arrangement of Fig. 8. Figure 14 shows the simple model used here to predict the zone of quiet created when the right-hand loudspeaker of the active headrest in Fig. 8 is modeled as a

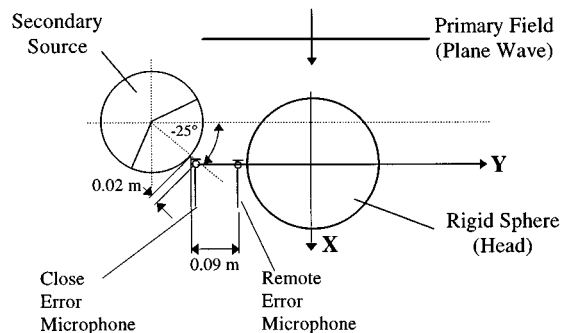


FIG. 14. The dual sphere model used to predict the zone of quiet created when the right-hand loudspeaker of Fig. 8 cancels the pressure at the right-hand error microphone in a single-channel local active noise control system.

spherical source of radius 0.08 m with an active segment of 120° , cancelling the pressure at a field point near a diffracting head, modeled as a rigid sphere of radius 0.115 m. The center line of the active segment forms an angle of -25° with the y axis as defined in Fig. 14. This angle corresponds to the drawing shown in Fig. 7.

For all the simulations reported here, the primary field is considered to be a plane wave propagating in the positive x direction as shown in Fig. 14, which corresponds to the experimental situation used in this paper. The two relative positions of the cancellation point theoretically studied are shown in Fig. 14. The close error microphone position represents a model of a “conventional” local active control system in which the error microphone is located close to the secondary source and at a noninterfering position so that the listener’s head can move freely. The remote error microphone position, on the other hand, depicts a model of an “acoustically ideal” case in which the error sensor is very close to an ear and remote from the secondary source to ensure a high attenuation at the ear position. Unfortunately, this error sensor position is impractical because it interferes with the head movements.

Figure 15 shows the calculated attenuation at the locations on the right-hand side (continuous) and left hand side (dashed line) of the diffracting sphere, when the secondary source is adjusted to cancel the primary pressure at the error microphone located at the close microphone position of Fig. 14, for different positions of the diffracting sphere along the x axis. The predictions shown correspond to four different values of the excitation frequency, i.e., 160, 320, 480, and 960 Hz. These frequency values were chosen to be consistent with the measurements reported above for the two-channel system. It should be emphasised that for each head position the secondary sound field has been re-calculated with the new distance between the spheres, the secondary source required to cancel the total pressure at the error microphone has been computed, and the sum of the primary and new secondary field at the ear positions computed. We observe that at 480 Hz and below, the attenuation at the closest ear position is only about 5 dB and reasonably independent of position, while almost no change is predicted at the other ear. The increase in the calculated attenuation at the closest ear position, at $f=960$ Hz, as the rigid sphere moves along the x direction is due to the diffraction effect produced by the rigid sphere on the secondary acoustic field. This diffraction effect is more important as the center of the sphere approaches the axis of the active segment and as the frequency increases.⁹

Figure 16 shows similar results but for a lateral displacement of the diffracting sphere along the y axis. In this figure, the shortest distance between the cancellation microphone and the rigid sphere is 2.5 cm which corresponds to $y = -7$ cm in the graphs. The displacements along the x and y axis are referred to the system of reference shown in Fig. 8(b). These results suggest that having an error microphone close to the secondary source and remote from the listener’s ear, gives very poor attenuation at the ear’s position.

Figure 17 shows the calculated attenuation at the right (continuous) and left hand (dashed line) ears when the sec-

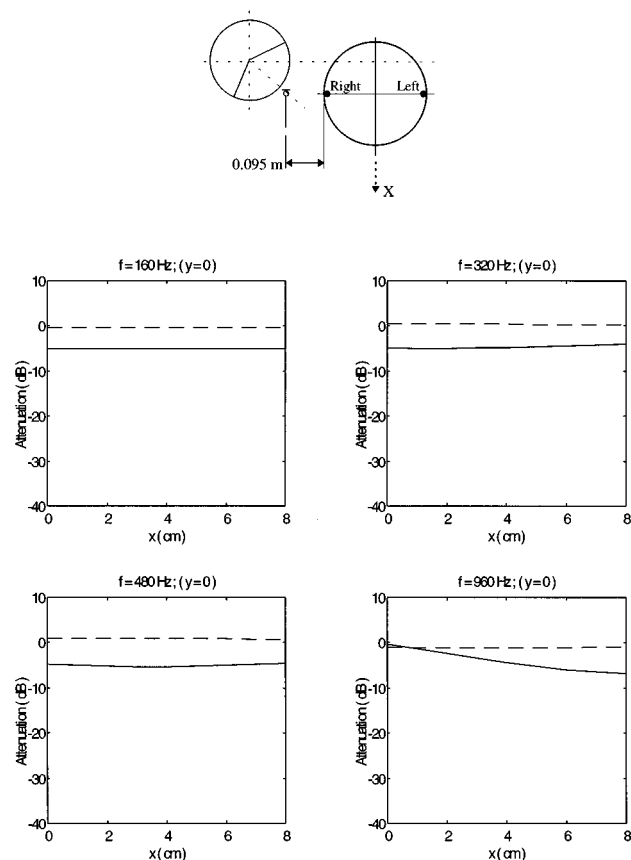


FIG. 15. The calculated controlled acoustic field at the measuring points corresponding to the right- (continuous) and left- (dashed line) hand ears of the listener (black dots), when the secondary spherical source cancels the primary acoustic field at the *close* error microphone location, for different positions of the rigid sphere along the x axis. The primary acoustic field is a plane wave propagating in the positive x direction.

ondary spherical source is adjusted to cancel the primary pressure at an error microphone located at the remote microphone position of Fig. 14, for different positions of the diffracting sphere along the x axis. Figure 18 shows the corresponding results for a lateral displacement of the diffracting sphere along the y axis. For this remote microphone arrangement, the attenuation that can be expected at the ear location is considerably higher than that obtained when the error sensor is close to the secondary source. This conclusion agrees with the experimental results presented above and, therefore, the model can be used to understand the physics and design other arrangements.

Note that because in the theoretical model the primary acoustic field is a plane wave parallel to the microphone arrangement, then the primary pressure is equal at the remote and virtual microphone positions. The theoretical model for the virtual microphone will thus produce results which are very similar to those produced by cancelling the pressure at the remote physical microphone (Figs. 17 and 18).

The acoustic pressure at the right and left ears of the manikin was measured after control with the close microphone arrangement of Fig. 14, with the right-hand loudspeaker adjusted to cancel the primary acoustic pressure at the right-hand error microphone. These experimental results were compared with the corresponding theoretical calcula-

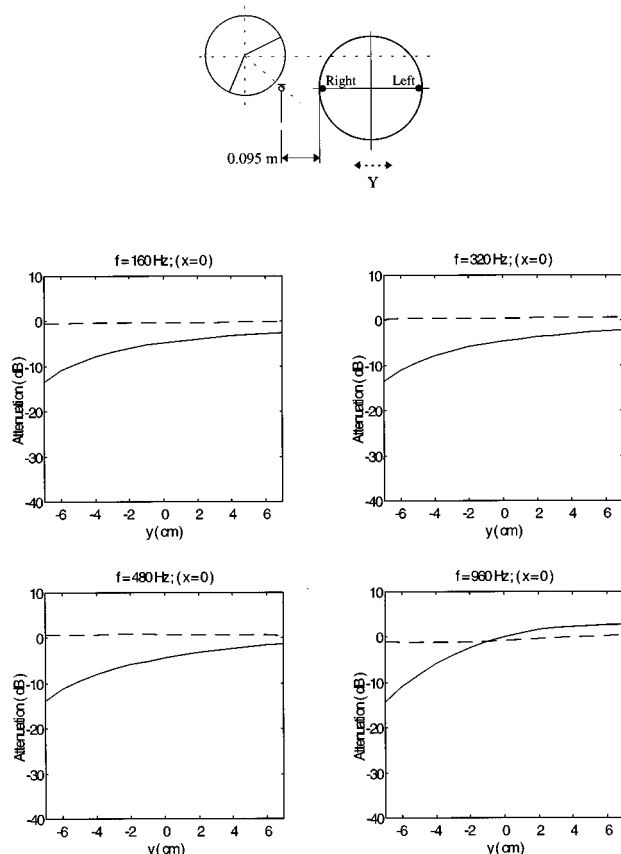


FIG. 16. The calculated controlled acoustic field at the measuring points corresponding to the right- (continuous) and left- (dashed line) hand ears of the listener (black dots), when the secondary spherical source cancels the primary acoustic field at the *close* error microphone location, for different positions of the rigid sphere along the y axis. The primary acoustic field is a plane wave propagating in the positive x direction.

tions shown in Figs. 15 and 16. The agreement found between the measurements and theoretical predictions is remarkably good up to a frequency of, about, 500 Hz. Even at frequencies as high as 960 Hz, the results of the simple theoretical model give a relatively good approximation to the experimental results. This agreement suggests that at low frequencies the zones of quiet generated by cancelling the pressure at a given point in the near field of a loudspeaker radiating near a diffracting head is mainly dictated by the relative dimensions and position of the secondary source and the diffracting head. We also observe that the two-sphere model gives good prediction of the two-channel results since the two channels behave almost independently.

The measured acoustic pressure after control for the remote microphone arrangement of Fig. 14 was also measured when the right hand loudspeaker is adjusted to cancel the primary pressure at the remote error microphone, for different positions of the manikin along the x and y axis. These results were compared with the theoretical predictions of Figs. 17 and 18, respectively. Again, a very good agreement between the experimental results and the theoretical predictions up to a frequency of 480 Hz was found. At higher frequencies, however, the disagreement between theory and measurements is slightly larger, presumably due to diffraction effects caused by the headrest and not considered in the

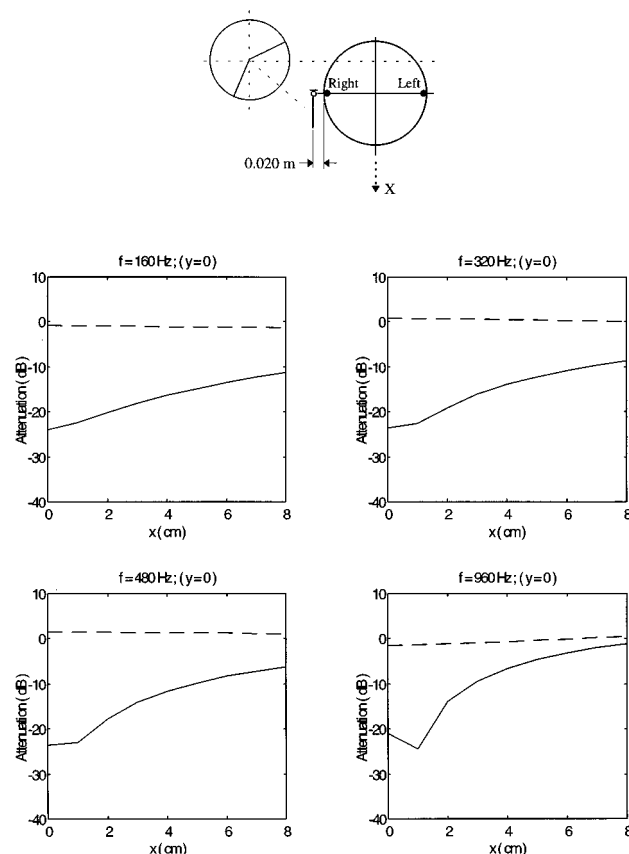


FIG. 17. The calculated controlled acoustic field at the measuring points corresponding to the right- (continuous) and left- (dashed line) hand ears of the listener (black dots), when the secondary spherical source cancels the primary acoustic field at the *remote* error microphone location, for different positions of the rigid sphere along the x axis. The primary acoustic field is a plane wave propagating in the positive x direction.

two-sphere model. In general, we can observe that the theoretical model used in these calculations, which does not take into account some of the other reflecting components of the headrest assembly, tends to predict slightly lower values of attenuation than those actually measured. This is consistent with the fact that an additional reflecting surface will induce a broadening of the zone of quiet created in the proximity of such a surface.¹²

The attenuation at the manikin's ears when the single channel virtual microphone arrangement in Sec. I is used was measured for different positions of the manikin in the x and y directions. It was found that the results for the attenuation at the right ear with the single channel system are almost identical to those shown in Figs. 12 and 13 for the two-channel system.

III. EFFECT OF A DIFFRACTING HEAD ON THE CONTROL ALGORITHM

If an adaptive feedforward controller is used to implement a local active control system, the output of the secondary acoustic sources are automatically adjusted in such a way as to minimize the sum of the squared error signals which are derived from a number of error microphones perhaps using Eq. (3). A reference signal which is at the same frequency as the primary excitation but is unaffected by the action of the

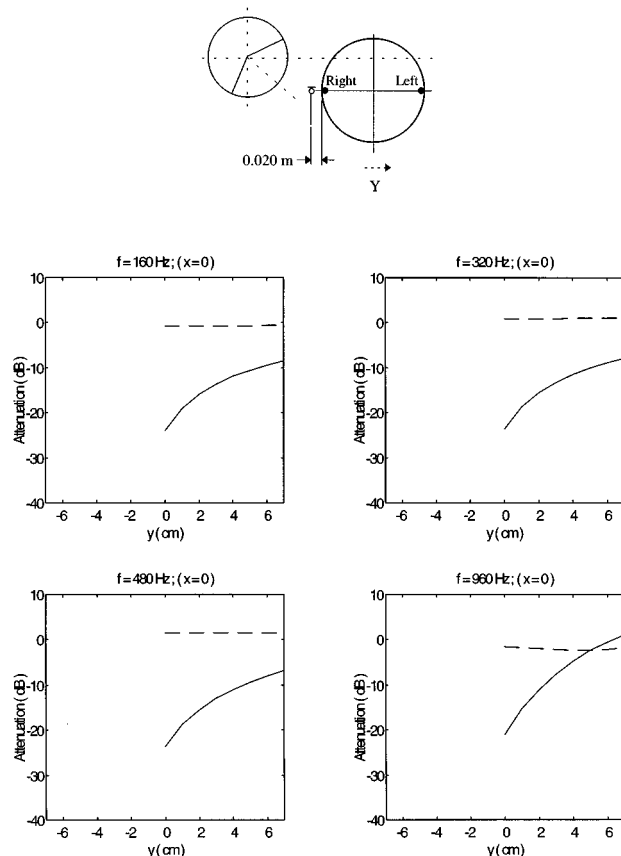


FIG. 18. The calculated controlled acoustic field at the measuring points corresponding to the right- (continuous) and left- (dashed line) hand ears of the listener (black dots), when the secondary spherical source cancels the primary acoustic field at the *remote* error microphone location, for different positions of the rigid sphere along the y axis. The primary acoustic field is a plane wave propagating in the positive x direction.

secondary sources is required. This reference signal is used to drive each of the secondary sources via an array of digital filters. The control problem is thus how best to adapt these digital filters to minimize the sum of the squared error signals. The LMS algorithm has shown itself to be very effective and robust when used in these situations. This algorithm, however, requires some estimate of the transfer response from each secondary source to each error sensor in order to generate the filtered reference signals needed to update the adaptive filter coefficients. These estimates correspond to a model of the acoustic “plant” being controlled and the accuracy of this model will affect the convergence of the adaptive algorithm but not affect its final steady-state behavior if the number of error sensors is equal to the number of secondary sources.

If, on the other hand, a fixed feedforward control is considered, the designer has to use the model of the plant and the measured primary field to calculate the fixed optimal secondary source strengths that achieve pressure cancellation at the desired acoustic field points. Since the system is not adaptive, any change in the primary field will cause a deterioration of the acoustic performance of the control system.

This section is concerned with the change in the transfer response from a secondary source to a physical and virtual microphone, produced by a diffracting head and the effect of

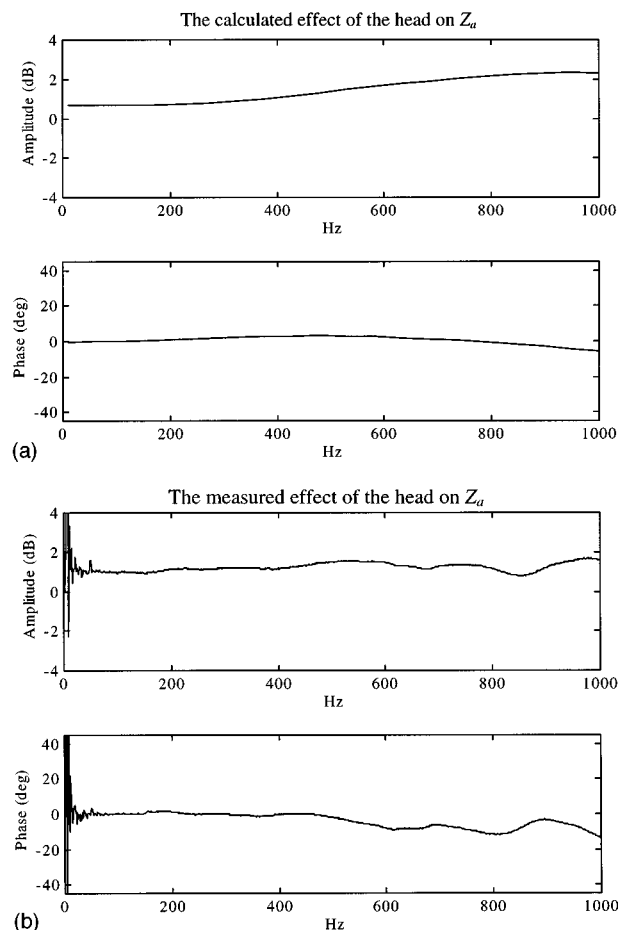


FIG. 19. (a) The *calculated* ratio between the transfer response from the right-hand loudspeaker to the *physical* microphone location with and without the diffracting sphere present. The location of the physical microphone corresponds to the close position shown in Fig. 14 and the center of the diffracting sphere is at $(x,y)=(0,-0.07)$ m. The secondary loudspeaker has been modeled as a spherical source of radius 0.08 m with an active segment of 120° whose center line forms an angle of -25° with the y -axis (see Fig. 14). The diameter of the diffracting sphere has been taken equal to 0.115 m. (b) The *measured* ratio between the transfer response from the right-hand loudspeaker of the practical system in Fig. 8 to the *physical* microphone location with and without the manikin present. The location of the physical microphone corresponds to the close position shown in Fig. 14 and the center of the manikin’s head was located at $(x,y)=(-0.07,0)$ m.

this change in the acoustic performance of a local active noise control system, using a physical and a virtual microphone, based on an adaptive controller.

A. Effect on the stability of an adaptive algorithm

The two-sphere model shown in Fig. 14 has been used to predict the change in the transfer response from the right hand loudspeaker to the physical and virtual microphone positions due to the presence of a diffracting head. As before, the position of the physical microphone corresponds to the close microphone arrangement of Fig. 14 and that of the virtual microphone, to the remote location depicted in the same figure. Figure 19(a) shows the calculated modulus and phase of the ratio between the acoustic transfer response from the spherical source to the physical microphone position, Z_a with and without the diffracting sphere present. The

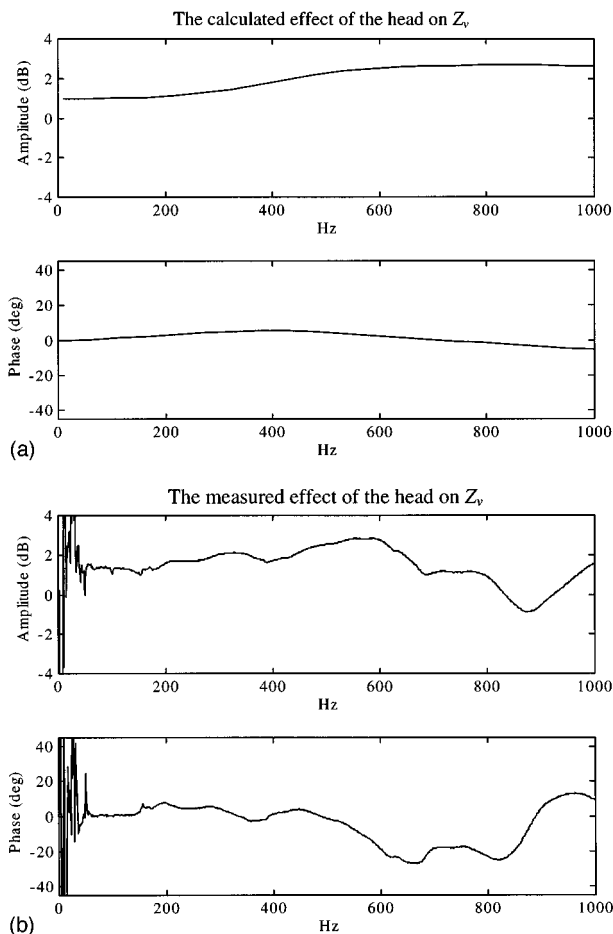


FIG. 20. (a) The *calculated* ratio between the transfer response from the right-hand loudspeaker to the *virtual* microphone location, Z_v with and without the diffracting sphere present. The location of the virtual microphone corresponds to the remote position shown in Fig. 14 and the center of the diffracting sphere is at $(x,y)=(0,0)$ m. The secondary loudspeaker has been modeled as a spherical source of radius 0.08 m with an active segment of 120° whose center line forms an angle of -25° with the y axis (see Fig. 14). The diameter of the diffracting sphere has been taken equal to 0.115 m. (b) The *measured* ratio between the transfer response from the right-hand loudspeaker of the system of Fig. 8 to the *virtual* microphone location with and without the manikin present. The location of the virtual microphone corresponds to the remote position in Fig. 14 and the center of the manikin's head was located at $(x,y)=(0,0)$ m.

position of the rigid sphere was taken equal to $(x,y)=(0,-0.07)$ m which corresponds to the closest position between the physical microphone and the head considered in the measurements reported in Sec II C. Figure 19(b) shows the measured change in Z_a produced by the manikin with respect to the case in which it is not present. In this measurement, the manikin was located so that the center of the head coincided with the position of the diffracting sphere, i.e., $(x,y)=(0,-0.07)$ m. The measurement was carried out in the same listening room used for the measurements of the zones of quiet reported above. Comparing Fig. 19(a) with 19(b), we observe that the two-sphere model gives a good prediction of the measured results.

Figure 20(a) shows the calculated modulus and phase of the ratio between the transfer response from the spherical source to the virtual microphone position, Z_v with and without the diffracting sphere present. In this case, the position of

the rigid sphere has been taken equal to $(x,y)=(0,0)$ which corresponds to the head laterally centered with respect to the headrest. Figure 20(b) depicts the corresponding measurements which agree well with the predictions.

These figures show that, up to a frequency of 1 kHz, the change in the transfer responses to the physical and virtual microphones due to the head movements is less than 3 dB in amplitude and less than $\pm 10^\circ$ in phase. We observe that the greatest discrepancy between the calculated and measured results corresponds to the change in Z_v , since the acoustic pressure at the virtual microphone position is affected more by reflections coming from surfaces other than the loudspeaker cabinet and the diffracting head than the pressure at the physical microphone location. From Figs. 19 and 20 we observe that the two-sphere model gives a good prediction of the measured results up to a frequency of about 600 Hz. Above this frequency, this agreement deteriorates slightly, probably due to the diffraction produced by the headrest and seat. Boucher *et al.*¹⁶ have shown that a single channel filtered-x LMS algorithm will converge, in the limit of slow adaptation, so long as the phase error in the plant model is less than 90° . They also show that at normal convergence rates the algorithm is relatively unaffected by phase errors in the plant model of less than approximately 45° . Therefore, the expected changes in the acoustic path of the local active control system considered here are small enough to ensure that an adaptive controller will converge to the optimal solution. However, if the controller was fixed, these changes would definitely worsen the acoustic performance of the control system.

B. Effect on the performance of an adaptive algorithm

In Sec. I we showed that a local active noise control system using a virtual microphone needs a model of the difference between the transfer responses from the secondary source to the physical and virtual microphones, i.e., $Z_a - Z_v$. The change in $Z_a - Z_v$ due to the presence of a head affects the estimate of the total pressure at the virtual microphone position which is minimized by the adaptive controller. Therefore, any differences between the value of $Z_a - Z_v$ assumed by the controller and its true value will cause a deterioration in the cancellation achieved at the virtual microphone position. The calculated ratio between the value of $Z_a - Z_v$ with and without the diffracting sphere present is shown in Fig. 21(a). The position of the diffracting head has been taken so that its center is located at $(x,y)=(0,0)$. The result shows that, up to about 1 kHz, the expected change in $Z_a - Z_v$ is less than 1 dB in amplitude and 10 degree in phase. We also observe that the value of $Z_a - Z_v$ with a diffracting sphere is about 1 dB less than when it is measured without a sphere present. This means that the secondary acoustic field in the points between the secondary source and the rigid sphere is spatially more uniform than the pressure distribution in free field, which is consistent with the observed increase in the extent of the zone of quiet when a diffracted head is present. Figure 21(b) shows the measured change in $Z_a - Z_v$ produced by the presence of the manikin with respect to the situation in which it is not present.

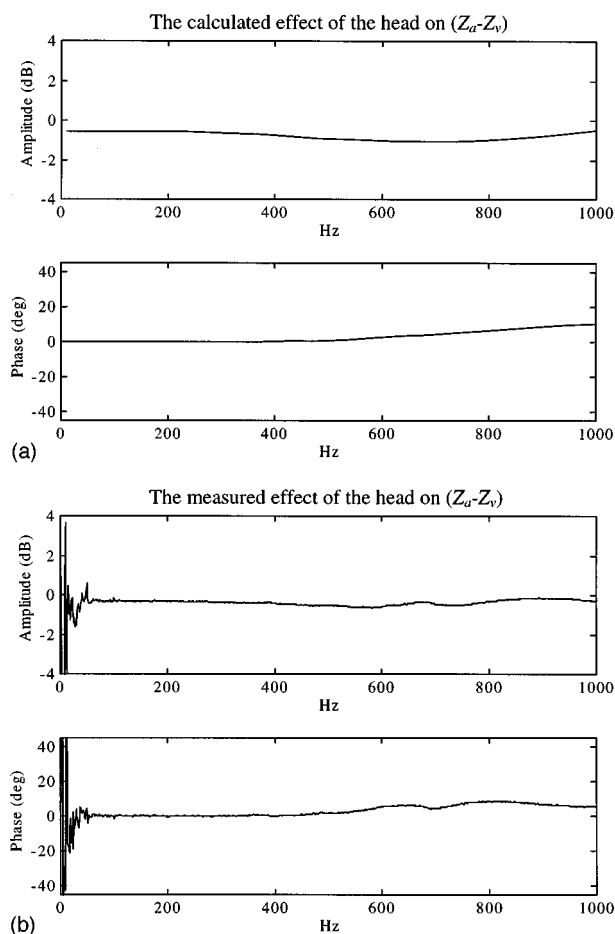


FIG. 21. (a) The *calculated* ratio between the difference in the transfer responses Z_a and Z_v i.e., $(Z_a - Z_v)$, with and without the diffracting sphere present. The position of the physical and virtual microphones correspond to the close and remote positions of Fig. 14, respectively, and the center of the diffracting sphere is at $(x, y) = (0, 0)$. The secondary loudspeaker has been modeled as a spherical source of radius 0.08 m with an active segment of 120° whose center line forms an angle of -25° with the y axis (see Fig. 14). The diameter of the diffracting sphere has been taken equal to 0.115 m. (b) The *measured* ratio between the difference in the transfer responses Z_a and Z_v , $(Z_a - Z_v)$, with and without a manikin present. The position of the physical and virtual microphone correspond to the closed and remote location in Fig. 14 respectively and the center of the diffracting sphere is at $(x, y) = (0, 0)$.

The previous results have shown the order of magnitude of the effect of a listener's head on the transfer responses from a loudspeaker to a physical and a virtual microphone locations in a practical local active control system. We are interested in estimating the effect that this change in the transfer responses have on the extension of the diffuse field zones of quiet produced by an adaptive controller using a virtual microphone arrangement.

Figure 22 shows the average diffuse field zones of quiet, for different frequencies or values of kL , created by a spherical secondary source with a 45-degree active segment and radius $L/2$ seeking to cancel the estimated pressure at a virtual microphone position, "*", located at $(x, y) = (L, 0)$ in the presence of a rigid sphere. The rigid sphere has a radius of $11L/15$ and its center is located at $(x, y) = (2L, 0)$. This configuration has been chosen to be consistent with other calculations carried out in Sec. I. The estimate of the total acoustic

pressure at the virtual microphone position, \hat{p}_v , is obtained after subtracting the term $q_s(Z_a - Z_v)$ from the total pressure measured by the physical microphone, where q_s is the secondary source strength, Z_a is the transfer response from the secondary source to the physical microphone position and Z_v is the transfer response to the virtual microphone location. The contour plots in Fig. 22 have been produced assuming that the adaptive controller uses an estimate of $Z_a - Z_v$ corresponding to the situation when the diffracting sphere is not present. Comparing Fig. 22 with Fig. 2(b) we observe that even though the change in $Z_a - Z_v$ due to the presence of the head is small, it has bad effects on the acoustic performance of the adaptive controller above $kL = 1$. These results suggest that in order to get good performance, i.e., attenuation at the virtual microphone position, then the difference $Z_a - Z_v$ must be measured with the head present. In the experiments shown in Sec. II, $Z_a - Z_v$ was measured *with* the head in the central position, therefore the change in $Z_a - Z_v$ caused by subsequent head motion are not important.

IV. CONCLUSIONS

A single virtual microphone arrangement that "projects" the zone of quiet further away from the secondary source than the position of the physical error sensor has been theoretically explored and its performance with and without the presence of a diffracting sphere has been assessed. The results suggest that this sort of arrangement is very sensitive to inaccuracies in the estimation of the primary acoustic component at the virtual microphone location from that at the physical microphone, especially at high frequencies.

The effect that the diffraction effects have on the attenuation at the virtual microphone position in a local active noise control system operating in a diffuse sound field have been theoretically studied. Analytical expressions for the average attenuation at a virtual microphone close to a rigid sphere, a wall, a two-wall edge, and a corner in a diffuse primary field have been derived. These results show that as the extension of the reflecting surface increases with respect to the acoustic wavelength, the primary diffuse field becomes more uniform and the performance of the virtual microphone arrangement improves considerably.

A practical local active noise control system has been built and installed in a headrest of a passenger seat. This arrangement has been used to measure the zones of quiet created by a single channel and a dual channel implementation of a local active noise control system near a listener's head. The experimental results for the single channel system have been compared with the predictions obtained using a two-sphere model which assumes that the secondary loudspeaker can be modeled as an active segment in a spherical cabinet and the listener's head is a rigid sphere. The agreement found between the predictions and the measured zones of quiet suggests that for frequencies below about 500 Hz, the zone of quiet in the near field of a secondary source close

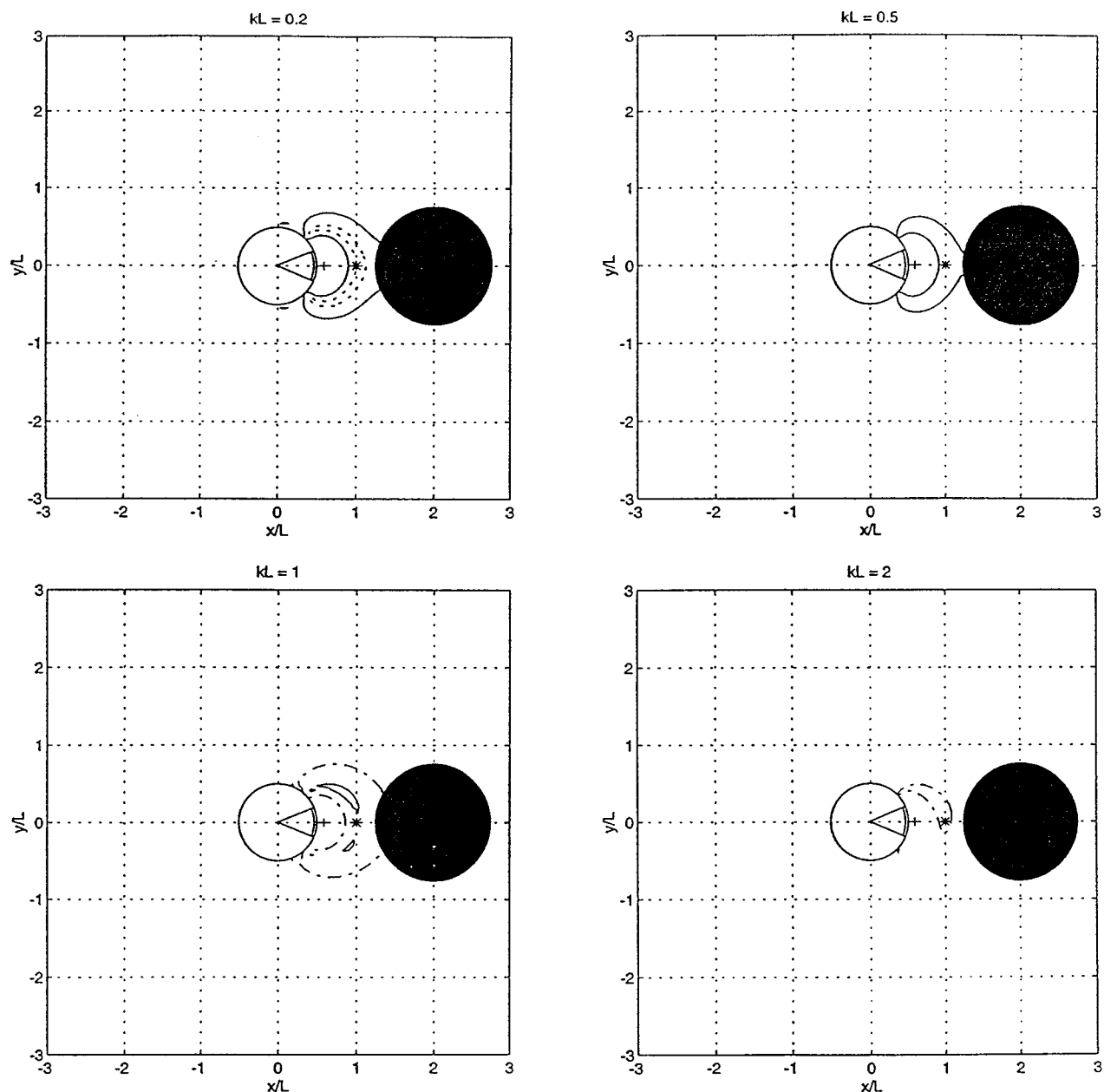


FIG. 22. The calculated average zone of quiet in the x - y plane due to the superposition of a primary diffuse sound field and the field due to a secondary spherical source with a 45° active segment and radius $L/2$ seeking to cancel the pressure at a *virtual* microphone (*) located at $(x,y)=(L,0)$ in the vicinity of a rigid sphere of radius $11L/15$ whose center is at $(x,y)=(2L,0)$. The physical microphone (+) is located at $(x,y)=(0.6L,0)$. The controller assumed is an *adaptive* feedforward controller with a model of the difference between the transfer responses Z_a and Z_v corresponding to the case of no rigid sphere present. (Dash-dot line) = -5 dB; (Continuous line) = -10 dB; (Dotted line) = -20 dB.

to a diffracting head is mainly dictated by the relative size and position of the loudspeaker secondary source, the listener's head, and the position of the cancellation point. The geometric details of the real headrest only produce a noticeable diffraction effect at frequencies near 1 kHz. For frequencies near 1 kHz, the two-sphere model does not contain sufficient geometric information to describe the diffraction effects caused on the secondary acoustic field. If more accurate predictions of the zone of quiet are required at these frequencies, a more detailed description of the geometric details of the headrest, head, and torso must be provided. This would also imply the need for more sophisticated numerical

methods based on FEM or BEM in order to describe the primary and secondary acoustic field near the listener's head.

A single channel and a dual channel system using virtual microphones located close to a manikin's head have also been experimentally investigated. The results show that a virtual microphone arrangement can effectively project the zones of quiet further away from the secondary loudspeaker than the position of the physical microphone. For the type of primary acoustic field used in the measurements, the performance of the local active control system using virtual and remote physical microphones is very similar for frequencies below about 500 Hz. We note that at 500 Hz, the distance

between the physical and virtual microphone is about 0.1λ . At frequencies of 1 kHz, however, the difference in the primary acoustic field at the physical and virtual microphone locations make the virtual microphone arrangement incapable of projecting the zones of quiet and, in some cases, even increases the total acoustic pressure after control. The results may have been different if other type of primary acoustic field had been used. If, for example, the primary acoustic field was a progressive plane wave moving in the y -direction (see Fig. 14) the virtual microphone implementation may have given a poorer performance at high frequencies than that reported here, due to the phase difference in the primary acoustic pressure at the physical and virtual microphone locations. It may thus be desirable to measure a spatially averaged zone of quiet for a number of different primary fields. However, this would involve the measurement of an ensemble of zones of quiet at different frequencies for many types of primary fields in order to average the results. This sort of measurement would then lead to the measurement of the diffuse field zone of quiet. At low frequencies, however, it is not anticipated that the diffuse results will be significantly different from those reported here.

The change in the acoustic transfer impedance from a secondary loudspeaker to the physical and virtual microphone positions when a diffracting head is moved close to the cancellation point have been calculated using the two-sphere model and measured with a realistic local active noise control arrangement. The observed changes in the transfer responses up to 1 kHz are small in both amplitude and phase which suggests that an adaptive LMS controller will be stable under these changes.

The effect of a diffracting head on the performance of a local active noise control system using an adaptive controller with a virtual microphone has been explored using the two-sphere model. The calculated diffuse field zones of quiet show that even though the change in $Z_a - Z_v$ due to the presence of the head is small, the effect of this change in the acoustic performance of an adaptive controller is important.

This implies that in order to get a good performance, i.e., attenuation at the virtual microphone, $Z_a - Z_v$ must be measured with the head present.

ACKNOWLEDGMENTS

This work was partially supported by "La Caixa" Foundation and the CIRIT (Generalitat de Catalunya), in Barcelona, Spain.

- ¹H. F. Olson and E. G. May, "Electronic sound absorber," *J. Acoust. Soc. Am.* **25**, 1130–1136 (1953).
- ²C. F. Ross, "Active control of sound," Ph.D. thesis, University of Cambridge, 1980.
- ³P. Joseph, S. J. Elliott, and P. A. Nelson, "Near field zones of quiet," *J. Sound Vib.* **172**(5), 605–627 (1994).
- ⁴P. Joseph, "Active control of high frequency enclosed sound fields," Ph.D. thesis, University of Southampton, 1990.
- ⁵A. David and S. J. Elliott, "Numerical studies of actively generated quiet zones," *Appl. Acoust.* **41**, 63–79 (1994).
- ⁶J. Garcia-Bonito and S. J. Elliott, "Local active control of diffracted diffuse sound fields," *J. Acoust. Soc. Am.* **98**, 1017–1024 (1995).
- ⁷S. J. Elliott and A. David, "A virtual microphone arrangement for local active sound control," 1st International Conference on Motion and Vibration Control, Yokohama, 1027–1031 (1992).
- ⁸P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, New York, 1992).
- ⁹J. Garcia-Bonito, "Local active control in pure tone diffracted diffuse sound fields," Ph.D. thesis, University of Southampton, 1996.
- ¹⁰S. J. Elliott, P. Joseph, A. J. Bullmore, and P. A. Nelson, "Active cancellation at a point in a pure tone diffuse sound field," *J. Sound Vib.* **120**(1), 183–189 (1988).
- ¹¹A. R. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York 1981).
- ¹²J. Garcia-Bonito, S. J. Elliott, and M. Bonilha, "Active cancellation of pressure at a point in a pure tone diffracted diffuse sound field," *J. Sound Vib.* **201**(1), 63–65 (1997).
- ¹³M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.* **58**(1), 214–222 (1995).
- ¹⁴S. Pheasant, *Body Space: Anthropometric Ergonomics and Design* (Taylor & Francis, London 1986), 1st ed.
- ¹⁵F. M. Wiener and D. A. Ross, "The pressure distribution in the auditory canal in a progressive sound field," *J. Acoust. Soc. Am.* **18**, 401–408 (1946).
- ¹⁶C. C. Boucher, S. J. Elliott, and P. A. Nelson, "Effect of errors in the plant model on the performance of algorithms for adaptive feedforward control," *IEE Proc. F* **138**(4), 313–319 (1991).

Characterization of a diffuse field in a reverberant room

H. Nélisse and J. Nicolas

G.A.U.S., Mechanical Engineering Department, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada

(Received 30 April 1996; revised 22 November 1996; accepted 31 December 1996)

An efficient modal approach to characterize the diffuseness of the sound field in a rectangular room is presented. Using two simple descriptors, the correlation function and the spatial uniformity of the pressure field, a practical and convenient tool is proposed to study the diffuse field in the room. A precise criterion has been given in terms of the least permissible number of room modes to achieve an adequate diffusion. It has been shown that the criterion is in great accordance with the well-known “Schroeder frequency” limit for the diffuse field. Detailed calculations of the correlation function are presented to show the importance of the $\Delta k/k$ correction terms to the well-known $\sin(kR)/kR$ prediction. A new closed form for the correlation function is thus derived. A discussion about the importance of having more than one descriptors is also presented. It is shown that a diffuse field can be established in a room with strong modal behavior under certain assumptions. © 1997 Acoustical Society of America. [S0001-4966(97)02105-X]

PACS numbers: 43.55.Br, 43.55.Cs, 43.55.Nd [JDQ]

INTRODUCTION

The degree of diffuseness of the acoustical field in a reverberant room has been widely investigated in the last five decades. However, even if the concept of diffuse field is well-understood, there are few practical and meaningful theoretical tools to characterize the sound field in a room. Moreover, many definitions of diffuseness and descriptors used to quantify the degree of diffuseness are often misused, if not misunderstood. As an example, the knowledge of only the uniformity of the pressure field in the room gives no sufficient information about the degree of diffuseness of the field. Generally, more than one descriptor have to be used to correctly define the sound field.

When dealing with reverberant test rooms, many questions can be raised by both experimentalists and theoreticians. For a given reverberant room, is it possible to create or simulate a diffuse field? If so, what are the frequency limits or what is the spatial extent of such a field? What are the basic assumptions that have been made to establish the diffuse field and what is the “quality” of the diffuseness? Since the room is used to conduct material testings, what will be the effect of these materials on the diffuseness? This paper tries to give some answers to these questions for an empty rectangular room, keeping in mind that the room is designed to contain testing structures. The main objective is to provide to acousticians concrete criterion of diffuseness using simple and well-known descriptors as well as a simple theoretical tool to compute these descriptors.

A simple analytical modal approach is presented to compute descriptors of interest in a rectangular reverberant room. Two indicators have been used: the correlation function and the spatial uniformity. The results are compared with the “perfect” diffuse field model, the so-called plane-wave model. The modal approach provides a useful tool to characterize the acoustic field in the empty chamber as well as in the one containing structures.

I. DEFINITION OF A DIFFUSE SOUND FIELD

A generally accepted definition of a diffuse field¹⁻³ can be given by: *An acoustic field is considered to be perfectly diffuse in a volume V if the energy density is the same on all points of this volume V.* It is possible to build such a field by a superposition of an infinite number of freely propagating plane waves, such that all directions of propagation are equally probable and the phase relations of the waves are random. This “construction” is usually called the plane-wave model (PWM). From a conceptual point of view this definition is quite adequate but it gives no practical help or criterion to determine the degree of diffuseness of an acoustical field.

II. METHODS TO CHARACTERIZE A DIFFUSE FIELD

Many experimental and theoretical methods have been proposed for evaluating diffuseness, each of these having their pros and cons. Brief reviews of various descriptors are presented by Schultz⁴ and more recently by Abdou.⁵ Bodlund⁶ has proposed an estimator ϵ_d based on the standard deviation of the correlation coefficients in regard to the theoretical predictions for a diffuse field. A stochastic model is used to compute the theoretical predictions. Property of the estimator ϵ_d has been experimentally verified. A complete statistical study of the diffuse field has been performed by Jacobsen⁷ using a similar stochastic model than the one used by Bodlund. Quantities such as distribution, mean and variance of the energy density or spatial autocovariance, and correlation function are studied. Comparisons between theoretical predictions and measurements are also presented.

Two of the most commonly used techniques are the cross-correlation and the spatial uniformity of the pressure field. These two have been chosen here as they can be both easily measured and, for simple rooms, easily computed.

Experimental works have been performed by many authors. Among all of these, Cook *et al.*⁸ have presented some results for the correlation function. Good agreements with

the $\sin(kR)/kR$ predictions are obtained for a high-frequency white-noise excitation. In the same spirit, Balachandran⁹ has presented one-third octave bands results for a reverberant room excited by a white noise. Once again, good agreements with the $\sin(kR)/kR$ form are found for high frequencies. Schroeder¹⁰ has computed, using an orthonormal modes expansion, the angular distribution of the energy density and some correlation functions for incident diffuse field on measuring walls. Pressure and pressure gradient are computed for both single frequencies and finite frequency bands. Recommendations and precautions are presented for the evaluation of the degree of diffuseness. Using the traversing microphone spectroscopy (TM), Lubman¹¹ has presented a technique to measure the autocorrelation function and the directivity function with an omnidirectional microphone for a 2-D and 3-D diffuse field. More recently, sound propagation and sound decay in reverberation rooms have been studied experimentally by Hodgson^{12,13} where results are compared with the method of images predictions. Zeng¹⁴ has computed the acoustic intensity and spatial distribution of the pressure field with a ray tracing approach. Comparisons with experimental results are also shown. It is shown that the field in the room is not perfectly diffuse, even for a high frequency excitation. Experimental results for the correlation function in a reverberation room has also been given by Koyasu and Yamashita¹⁵ as well as results for the directional pattern in the room. It is pointed out that it is important to observe the correlation coefficients for all directions in the field. Correlation coefficients along a line and directivity power spectra have been both computed and measured in a rectangular room by Tohyama *et al.*^{16,17} Good agreements between measurements and theoretical predictions are found but no quantitative assessment of sound field has been proposed.

On the theoretical side, Morrow¹⁸ has computed the correlation function in a cavity. A high-modal density is assumed and cross terms in the correlation function are neglected. Blake and Waterhouse¹⁹ have calculated the correlation function for an isotropic and anisotropic diffuse field. It is shown that anisotropy has no effect on the real part of the correlation but strongly modifies the imaginary part. Chien and Soroka²⁰ have calculated the correlation function in a high-modal cavity at high frequency for a stationary and a decaying state. In the stationary case, the $\sin(kR)/kR$ predictions are obtained. These calculations, as well as those of Morrow,¹⁸ for the correlation function are discussed by Chu.^{21,22} A discussion on the cross terms for the correlation function is presented. Computations of the spatial uniformity have been performed by Waterhouse²³ and Chu.²⁴ Their main conclusions are that spatial uniformity cannot be established for a monochromatic source. More recently, Kubota and Dowell²⁵ have proposed an asymptotic modal analysis (AMA) to study the spatial uniformity at high frequencies in a cavity. It is shown that AMA gives better results than the ray tracing technique.

This brief review reveals that, while a great amount of work has been done, no simple and clear criterion have been stated in terms of modal density. Generally speaking, a ‘‘Schroeder frequency’’-type limit is given but it gives no

clear insight on the number of modes per band required to achieve an adequate diffusion. Most of the studies have been dedicated to high frequencies and there is no precise indication on how to characterize the acoustic field when strong modal behavior is involved, i.e., at lower frequencies. In this sense, this paper is addressed to the study of cavity and frequency range for which modal behavior is important. Among all the descriptors commonly proposed to characterize the sound field, two of them have retained our attention in this paper: the correlation function and the spatial uniformity. These two descriptors are reliable, easy to measure and easy to compute in the framework of the PWM model and the modal approach.

A. Correlation function

The spatial cross-correlation function of the pressure field between two points, or simply the correlation function, is defined by

$$C(\mathbf{r}, \mathbf{r} + \mathbf{r}') = \frac{\overline{P(\mathbf{r})P(\mathbf{r} + \mathbf{r}')}}{\sqrt{\overline{P^2(\mathbf{r})}\overline{P^2(\mathbf{r} + \mathbf{r}')}}}. \quad (1)$$

The acoustic pressure at \mathbf{r} is given by $P(\mathbf{r})$ while the horizontal bar means time average.

It will be shown later that the plane-wave model leads to

$$C(\mathbf{r}, \mathbf{r} + \mathbf{r}') = \frac{\sin(kR)}{kR}, \quad (2)$$

with $R = |\mathbf{r} - \mathbf{r}'|$.

This function has been widely studied theoretically^{18,23,19,21,20} and experimentally.^{8,9,15–17} Generally, the plane-wave prediction $\sin(kR)/kR$ is obtained if a narrow-band containing enough chamber modes is used. In most of the theoretical studies, certain assumptions have been made to simplify the calculations of the correlation function: (i) in the modal expression of the correlation function, cross terms are neglected using a high-frequency approximation; (ii) high-frequency approximation allows to change the discrete summation over chamber modes to a continuous frequency integration. This assumption is valid only if the modal density is high; (iii) Detailed calculation of the correlation function shows that a $\Delta k/k$ correction applies to the well-known $\sin(kR)/kR$ form but it is always neglected.

One of the goals in this paper is to compute the correlation function without approximations, in order to extend the previous calculations to lower frequencies range and to take into account cross terms in the modal expansions.

B. Spatial uniformity

To study the spatial uniformity of the pressure field in a volume V , the standard deviation σ of the acoustic pressure field, which is related to the variance, is used. If one considers N points for which the sound-pressure levels $\text{SPL}(\mathbf{r})$ are computed in the central volume of the room, one defines the standard deviation to be given by

$$\sigma^2(f) = \frac{1}{N-1} \sum_{i=1}^N [\text{SPL}(\mathbf{r}_i, f) - \overline{\text{SPL}(f)}]^2, \quad (3)$$

TABLE I. Maximum admissible standard deviation for third-octave band as prescribed by the ISO standard #3741.

| Central frequency (one-third-octave band) | Maximum standard deviation (dB) |
|--|------------------------------------|
| 100–160 | 1.5 |
| 200–630 | 1.0 |
| 800–2500 | 0.5 |
| 3150–10 000 | 1.0 |

where $\overline{\text{SPL}}(f)$ is the mean sound-pressure level in the volume V at the frequency f . At the point \mathbf{r}_i in V , the sound-pressure level $\text{SPL}(\mathbf{r}_i, f)$ is defined by

$$\text{SPL}(\mathbf{r}_i, f) = 10 \log(|P(\mathbf{r}_i, f)|^2). \quad (4)$$

It has been accepted that an acoustic field in a qualified reverberation room exhibits adequate diffuseness if the standard deviation σ remains under 1.5 dB. To be more precise, the ISO standard²⁶ states that the maximum admissible standard deviation must not exceed the values given in Table I. Naturally, the plane-wave model gives $\sigma=0$ dB.

C. Modal approach

To compute the acoustic pressure generated in an empty reverberant room that exhibits strong modal behavior, a modal expansion model, noted MODAP, has been used. A rigid-walled rectangular room (dimensions $L_x \times L_y \times L_z$) is considered, as shown in Fig. 1. The room can be excited by a vibrating surface S_s lying in the $z=0$ plane. It is to be noted that only one acoustic source has been used to help better understand the physical phenomena. However, the present model can be easily adapted to take into account many sources. The wall surfaces are considered to be perfectly reflective. The acoustic pressure in the room can be computed with a modal approach that uses a two-indices empty cavity Green's function.^{27,28} It allows to reduce the number of summations by one which is a great improvement for algorithm efficiency. Appendix A draws a summary of this modal approach. With this expansion the acoustic pressure reads

$$P(\mathbf{r}, f) = \sum_{mn} P_{mn}(z, f) \psi_{mn}(x, y) \quad (5)$$

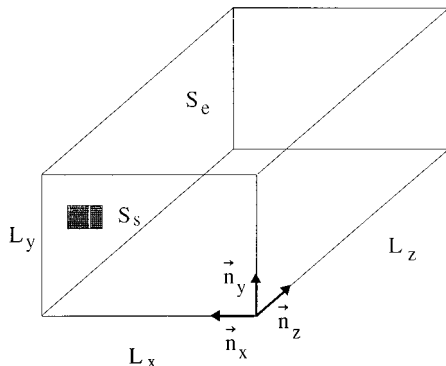


FIG. 1. Geometrical representation of an empty rectangular room excited acoustically.

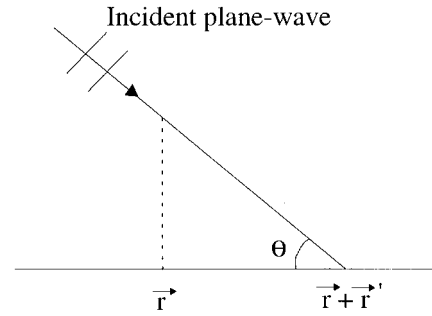


FIG. 2. Representation of an incident plane wave.

where the coefficients $P_{mn}(z, f)$ and the basis functions $\psi_{mn}(x, y)$ are given in Appendix A.

To include losses in the fluid, an acoustic damping factor η is used. It relates the wave number k to the piston driving frequency ω by the relation

$$k^2 = \frac{\omega^2}{c^2} \frac{1}{(1 + j\eta)}. \quad (6)$$

The damping factor η can be related to the reverberation time of the room if an exponential decay is assumed after the source has been cut off.

The correlation function and the spatial uniformity has been computed with this model and compared with the PWM predictions. The results are presented in Sec. IV.

III. PWM PREDICTIONS FOR THE CORRELATION FUNCTION

One wishes to compute the correlation function [Eq. (1)] for the plane-wave model. Consider the points \mathbf{r} and $\mathbf{r} + \mathbf{r}'$ as shown in Fig. 2. The pressure at these points is given by

$$P(\mathbf{r}) = A e^{i\mathbf{k} \cdot \mathbf{r}} \quad \text{and} \quad P(\mathbf{r} + \mathbf{r}') = A e^{i\mathbf{k} \cdot (\mathbf{r} + \mathbf{r}')}. \quad (7)$$

If a harmonic state is assumed for the pressure field, the correlation function becomes

$$\begin{aligned} C(kR) &= \text{Re} \left(\frac{P(\mathbf{r}) P^*(\mathbf{r} + \mathbf{r}')}{\sqrt{|P(\mathbf{r})|^2 |P(\mathbf{r} + \mathbf{r}')|^2}} \right) \\ &= \text{Re}(e^{-ikR \cos \theta}) = \cos(kR \cos \theta). \end{aligned} \quad (8)$$

It is to be noted that the correlation function only depends on the distance R . For a diffuse field an integration over all the possible directions (θ, ϕ) of the plane waves has to be performed. It gives the following results for 1, 2, and 3 dimensions:

1D

$$C(kR) = \cos(kR). \quad (9)$$

2D

$$C(kR) = \frac{1}{2\pi} \int_0^{2\pi} \cos(kR \cos \theta) d\theta = J_0(kR), \quad (10)$$

where $J_0(kR)$ is the first-order Bessel function of the first kind.

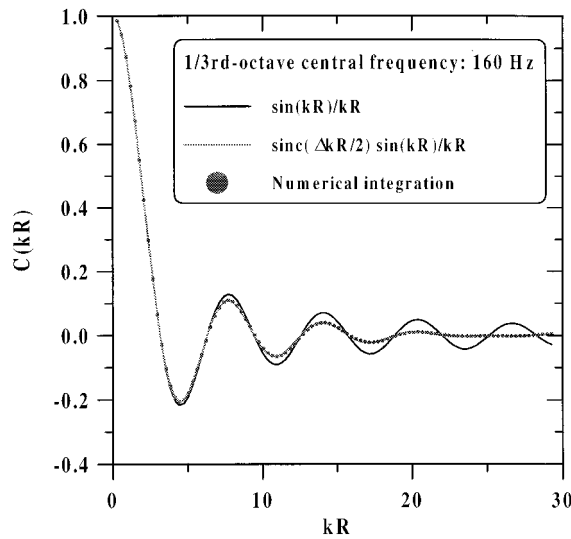


FIG. 3. Correlation function $\sin(kR)/kR$ (bold line), $C(kR)$ from Eq. (13) (gray line), and from the numerical integration of Eq. (11) (gray circles) as a function of kR .

3D

$$C(kR) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \cos(kR \cos \theta) \sin \theta \, d\theta \, d\phi$$

$$= \frac{\sin(kR)}{kR}. \quad (11)$$

If one is interested in a frequency-band response, one must integrate the correlation function over the frequency in the bandwidth $\Delta f = f_2 - f_1$. Appendix B gives the calculation of the correlation function in the 3D case. It leads to:

$$C_{\Delta f}(kR) \sim \text{sinc}\left(\frac{\Delta kR}{2}\right) \left[\frac{\sin(kR)}{kR} - 2 \frac{\cos(kR)}{(kR)^2} \sin^2\left(\frac{\Delta kR}{4}\right) \right], \quad (12)$$

where k is now defined by $k = (k_2 + k_1)/2$ and where $\text{sinc}(x) = \sin(x)/x$.

The above form for $C_{\Delta f}(kR)$ is quite different than those one can find in the literature for which only additional correction terms are found from a Taylor expansion. It is to be noted that for almost all practical situations, the second term of (12) can be neglected so that Eq. (12) can be written:

$$C_{\Delta f}(kR) \sim \text{sinc}\left(\frac{\Delta kR}{2}\right) \frac{\sin(kR)}{kR}, \quad (13)$$

It then gives a simple closed form for the influence of the $\Delta k/k$ correction.

To validate the last relation for $C_{\Delta f}(kR)$ [Eq. (13)], the frequency integrated equation (11) has been computed numerically using a simple Simpson's rule algorithm for the wave number integral. Figure 3 presents the results from the simple form $\sin(kR)/kR$ (bold line), from Eq. (13) (gray line) and from the numerical integration of Eq. (11) (gray circles) as a function of kR . The figure shows that Eq. (13) is a very good approximation for the finite frequency band correlation

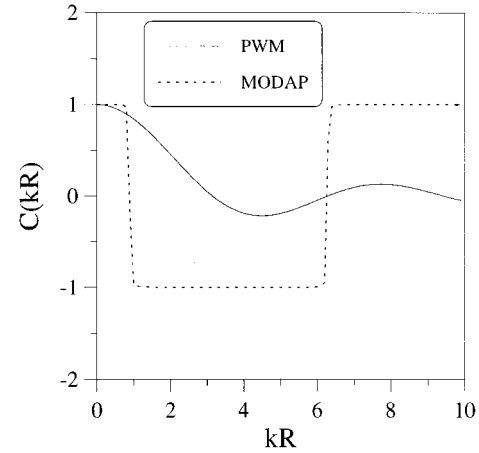


FIG. 4. Correlation function for the 121.6–121.7 Hz band.

function. The importance of the $\Delta k/k$ correction to the usual $\sin(kR)/kR$ form clearly appears for high values of kR .

IV. RESULTS

A. Correlation results

To compute the correlation with the MODAP model, one has used

$$C(kR)_{\Delta f} = \frac{\text{Re} \sum_{f'=f_1}^{f_2} P(\mathbf{r}) P^*(\mathbf{r}+\mathbf{r}')}{\sqrt{\sum_{f'=f_1}^{f_2} |P(\mathbf{r})|^2 \sum_{f'=f_1}^{f_2} |P(\mathbf{r}+\mathbf{r}')|^2}}, \quad (14)$$

where the acoustic pressures is given by Eq. (5).

If there is only one chamber mode in the bandwidth or for a monochromatic source, one can show that the correlation function becomes 0 or ± 1 . This fact is well illustrated in Fig. 4. A $8.0 \times 6.9 \times 9.75 \text{ m}^3$ chamber with a damping $\eta = 5 \times 10^{-3}$ has been used to compute the correlation function for the 121.6–121.7 Hz frequency band as a function of kR . For this room, only the (3,3,4) mode has its natural frequency in the frequency band of interest ($f_{334} = 121.65 \text{ Hz}$). All the different points used to compute the correlation have been chosen in order to avoid the effect of the walls on the acoustic pressure.^{18,24} The damping has been chosen in order to simulate losses in real cavity.²⁴ The dashed line represents the MODAP predictions while the solid line gives the PWM predictions. A value of ± 1 is correctly obtained for the correlation function with the MODAP technique. Similar results have been obtained for higher frequencies. It suggests that a diffuse field cannot be established in a room for a monochromatic source, a result not predicted by the plane-wave model. Chu²² has presented similar experimental results for which the same conclusion has been drawn.

Figure 5 shows the correlation function for the same chamber and damping with, now, a frequency band ranging from $f_1 = 90 \text{ Hz}$ to $f_2 = 112 \text{ Hz}$. For the test room used here, there exist 49 modes in this bandwidth. Three cases have been shown: the MODAP predictions and the PWM with and without the $\Delta k/k$ correction. One can see that the MODAP model gives satisfactory results, in particular for low values

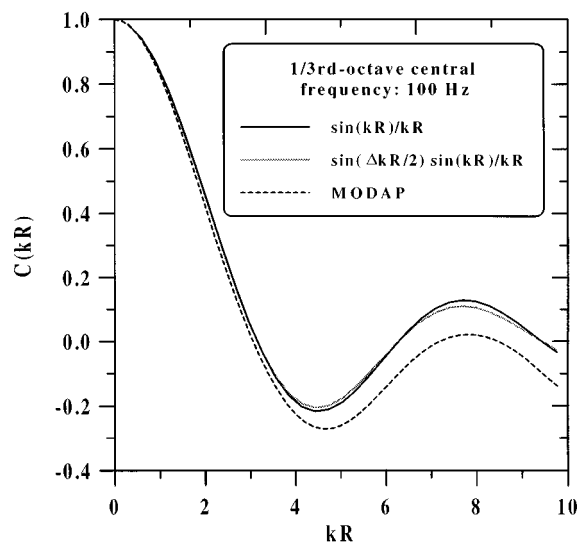


FIG. 5. Correlation function for the 90–112 Hz band.

of kR . Above $kR=4$, discrepancies between MODAP and PWM are observed. It suggests that the field is not perfectly diffuse (in the sense of the PWM). However, since the MODAP predictions are far from the single-frequency results discussed above [$C(kR) = \pm 1$], it is believed that these results indicate that a diffuse field can be obtained for a measurement bandwidth containing enough chamber modes. By the way, it gives no indication on the least permissible number of room modes to achieve adequate diffuseness. To do so, one can use the spatial uniformity as an indicator. Moreover, as pointed out by Schultz,⁴ the correlation function is not very sensitive to deviations from perfect diffuseness.

B. Spatial uniformity results

To study the spatial uniformity of the sound field, the relations (3) and (4) have been used. If one is interested in a frequency-band response one has:

$$\text{SPL}(\mathbf{r}, f) = 10 \log \left(\sum_{f'=f_1}^{f_2} |P(\mathbf{r}, f')|^2 \delta f' \right), \quad (15)$$

with $f = (f_2 + f_1)/2$, the center of the band.

A number of 20 points have been chosen in a volume V far from the walls (at least $\lambda/2$ from the nearest wall).²⁴ The indicator $\sigma(f)$ is shown in Fig. 6 for the $8.0 \times 6.9 \times 9.75 \text{ m}^3$ chamber with a damping $\eta = 5 \times 10^{-3}$. One-third octave bands have been used. It has been verified that a sufficient number of frequency points has been chosen to ensure convergence. In the same way, Table II gives the number of room modes in each of the one-third-octave band used for the computation of $\sigma(f)$. It is observed that the value $\sigma(f)$ begins to be in accordance with the prescribed values of Table I, less than 1.5 dB, at the 90–112 Hz band. For this band, one has $\sigma(f) \sim 1.3 \text{ dB}$, which is in accordance with the previous results for the correlation function (Fig. 5). From these results, a tentative criterion can be stated: A diffuse field can be established in a rectangular

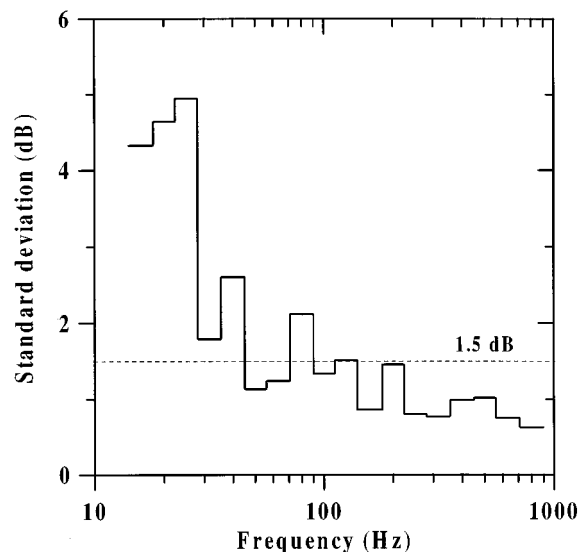


FIG. 6. Descriptor $\sigma(f)$ as a function of the frequency for one-third-octave bands.

room if there is at least 20–30 modes in the measurement bandwidth. For a bandwidth which contains less modes there is no adequate diffuseness, in particular at a given frequency (monochromatic source). For lower dimensions (1D and 2D rooms), different values have been obtained: 2 modes/band at 1D and 6 modes/band at 2D.

Similar simulations have been performed for different room dimensions to examine if a possible “critical” frequency above which the field is assumed diffuse can be estimated (analogous to the “Schroeder frequency”). Figure 7 shows the standard deviation for three rooms with different volumes using the MODAP as a function of kL where L

TABLE II. Number of modes in the one-third-octave bands for the $8.0 \times 6.9 \times 9.75 \text{ m}^3$ room.

| $f_1 - f_2$ | Number of modes/band |
|-------------|----------------------|
| 5.6–7.1 | 0 |
| 7.1–9 | 0 |
| 9–11.2 | 0 |
| 11.2–14 | 0 |
| 14–18 | 1 |
| 18–22.4 | 1 |
| 22.4–28 | 2 |
| 28–35.5 | 3 |
| 35.5–45 | 4 |
| 45–56 | 9 |
| 56–71 | 14 |
| 71–90 | 32 |
| 90–112 | 49 |
| 112–140 | 95 |
| 140–180 | 206 |
| 180–224 | 347 |
| 224–280 | 678 |
| 280–355 | 1380 |
| 355–450 | 2773 |
| 450–560 | 4975 |
| 560–710 | 10617 |
| 710–900 | 21371 |
| 900–1120 | 38612 |

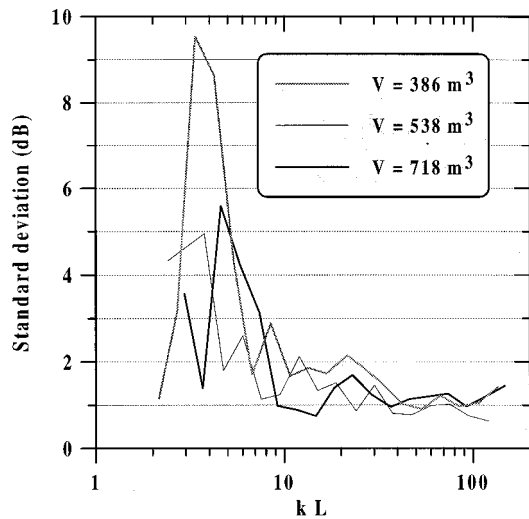


FIG. 7. Descriptor $\sigma(f)$ as a function of the frequency for three different rooms' volumes.

$= \sqrt[3]{V}$ is a characteristic dimension of the room. If one approximates the critical wave number $k_c = 2\pi f_c/c$ to be the one for which the standard deviation remains stable and under 2 dB, one has

$$k_c \approx \frac{22}{\sqrt[3]{V}}. \quad (16)$$

which gives, with $c = 340$ m/s

$$f_c \approx \frac{1190}{\sqrt[3]{V}}. \quad (17)$$

For the room with $V = 538 \text{ m}^3$ it gives $f_c = 146$ Hz.

From our damping model [see Eq. (6)] it can be shown that the “Schroeder frequency”,²⁹ is given by

$$f_s \approx \sqrt[3]{\frac{\alpha c^3}{4\pi\eta V}}, \quad (18)$$

where α is the modal overlap. Schroeder has proposed a modal overlap $\alpha = 3$. For a damping of $\eta = 5 \times 10^{-3}$ one then obtains $f_s = 152$ Hz. This is in great accordance with the “critical” frequency f_c found above. Meanwhile, it has to be noted that the choice of the factor 22 in (16) (and by the way the modal overlap of 3) is a quite subjective one. However, it gives a good approximation for the “critical” frequency and shows how the present model can be related to the well-known “Schroeder frequency.”

From the above results it becomes clear that a frequency limit for which the diffuse field exists can be estimated under certain conditions. In the limit where there is enough modes in the measurements bands the usual “Schroeder frequency” relation can be applied to estimate the diffuse field frequency limit. For very narrow bands the frequency limit will be higher than for broadband ones and no clear relation can be given for this “critical frequency.” This frequency depends on the room dimensions, the frequency, and the width of the band. With this information, the spatial uniformity is a practical tool to determine from which band the field begins to be

considered as diffuse, in terms of modal density. Meanwhile, the use of the correlation function assures whether or not if the field really presents adequate diffuseness. It is thought that the two descriptors must be used together to insure the quality of the diffuseness. The plane-wave model gives a good example of a case where only the use of the spatial uniformity is not sufficient to characterize a sound field.

V. CONCLUSION

The present paper has been dedicated to the study of the characterization of the sound field in a rectangular cavity. Even if the concept of diffuse field is well-understood, thanks to the numerous papers on this topic, there is a need for practical tools and in-depth analysis of diffuseness in the sense of the plane-wave model in relation with field that exhibits strong modal behavior.

By the use of two descriptors, the correlation function and the spatial uniformity, this paper has presented a simple and convenient model to characterize the sound field in a rectangular cavity. The main advantage of the proposed approach is the ability to simply implement it, both experimentally and analytically. The use of a two-indices Green's function allows to obtain a fast computation code. Emphasis has been put on the fact that only one descriptor is not sufficient to define correctly the diffuse field. Results have shown that a diffuse field can be established in a volume V if there is at least 20–30 room modes in the measurement band. This condition applies for all types of frequency bands. Relation between the present model and the well-known “Schroeder frequency” has been discussed. Moreover, a computation of the correlation function in the framework of the plane-wave model has been presented for 1, 2, and 3 dimensions. The importance of the $\Delta k/k$ correction to the generally used form for the correlation function has been discussed. A new simple closed form for the finite-frequency band correlation function has been given. It has been validated by the modal approach.

An important conclusion of the paper lies in the fact that a diffuse field can be established in a room that exhibits strong modal behavior. Even if it seems simple, this assertion indicates that an acoustic field without modal characteristics can be simulated by an a priori different approach with less computation costs. It also validates the generally used approximation which consists in neglecting the walls effects on the pressure by generating the diffuse field with a infinite number of plane waves. The present model allows to give the limitations of the plane-wave model with precise criterion.

While being a good starting basis, the present approach would benefit from experimental validations. It would help experimentalists to conduct practical works with insightful indications such as orientation and location of microphones in the room, spatial and/or frequency averaging, and so on. In this sense, a parametric analysis of the present model, with the help of experimental results, could provide this type of information.

ACKNOWLEDGMENTS

The authors would like to thanks the reviewers of this paper for very helpful suggestions, in particular for the computation of the correlation function. This work was supported by grants from the National Sciences and Engineering Research Council of Canada (NSERC).

APPENDIX A: MODAL-CAVITY APPROACH FOR THE ACOUSTIC PRESSURE

One has to compute the acoustic pressure in a rectangular cavity. Consider first the cavity shown in Fig. 1. The acoustic source is given by a vibrating surface S_s while the surface of the walls is noted S_e . The ‘‘Kirchoff–Helmholtz’’ equation for this problem reads:

$$P(\mathbf{r}_0) = \int \int_{S_s} dS G(\mathbf{r}, \mathbf{r}_0) \nabla_{\mathbf{r}} P(\mathbf{r}) \cdot \mathbf{n}_s \quad (\text{A1})$$

where the walls have been considered to be perfectly rigid. This condition can be written

$$\nabla_{\mathbf{r}} G(\mathbf{r}, \mathbf{r}_0) \cdot \mathbf{n}_e = 0 \quad \forall \mathbf{r} \in S_e. \quad (\text{A2})$$

Instead of expanding the Green’s function on cavity modes, which implies a triple summation, the function is expanded on a set of two-dimensional functions, reducing the number of summations by one. Such an expansion is written, as shown by Bruneau,²⁷ in the following form:

$$G(\mathbf{r}, \mathbf{r}_0) = \sum_{tu} g_{tu}(z, z_0) \psi_{tu}(x, y) \psi_{tu}(x_0, y_0). \quad (\text{A3})$$

The orthonormal basis functions ψ_{tu} are given by

$$\psi_{tu}(x, y) = \frac{4}{(1 + \delta_{t0})(1 + \delta_{u0})L_x L_y} \cos\left(\frac{t\pi x}{L_x}\right) \cos\left(\frac{u\pi y}{L_y}\right), \quad (\text{A4})$$

where δ_{tu} is the Kronecker delta and the function $g_{tu}(z, z_0)$ is

$$g_{tu}(z, z_0) = \begin{cases} -\frac{\cos[k_{ztu}(L_z - z_0)] \cos(k_{ztu}z)}{k_{ztu} \sin(k_{ztu}L_z)}, & \text{if } z \leq z_0, \\ -\frac{\cos[k_{ztu}(L_z - z)] \cos(k_{ztu}z_0)}{k_{ztu} \sin(k_{ztu}L_z)}, & \text{if } z \geq z_0, \end{cases} \quad (\text{A5})$$

where

$$k_{ztu}^2 = k^2 - \left[\left(\frac{t\pi}{L_x} \right)^2 + \left(\frac{u\pi}{L_y} \right)^2 \right]. \quad (\text{A6})$$

With this Green’s function, all the modes along the z axis are implicitly present so faster convergence can be achieved comparing to the classical three-indices Green’s function.

If one considers that the surface S_s vibrates harmonically with the frequency ω and with the velocity V_ω , one has:

$$P(\mathbf{r}) = i\omega\rho V_\omega \int \int_{S_s} dS G(\mathbf{r}, \mathbf{r}_0). \quad (\text{A7})$$

Using the above Green’s function (A3), the acoustic pressure in the room reads

$$P(\mathbf{r}) = -i\omega\rho \sum_{tu} \frac{V_{tu} \cos[k_{ztu}(L_z - z)]}{k_{ztu} \sin(k_{ztu}L_z)} \psi_{tu}(x, y). \quad (\text{A8})$$

The term V_{tu} is obtained from the integration over the piston surface and is given by

$$V_{tu} = V_\omega \int_{S_s} dS \psi_{tu}(x, y). \quad (\text{A9})$$

APPENDIX B: CALCULATION OF THE CORRELATION FUNCTION

One wants to compute the correlation function $G_{\Delta f}(kR)$ given by

$$C_{\Delta f}(kR) = \frac{1}{\Delta k} \int_{k_1}^{k_2} \frac{\sin(kR)}{kR} dk. \quad (\text{B1})$$

By a simple change of variables, the integral can be rewritten

$$C_{\Delta f}(kR) = \frac{1}{\Delta k} \int_{-\Delta k/2}^{\Delta k/2} \frac{\sin[(k + k_c)R]}{(k + k_c)R} dk, \quad (\text{B2})$$

where $k_c = (k_1 + k_2)/2$ and $\Delta k = k_2 - k_1$.

One first uses the following trigonometric identity

$$\sin[(k + k_c)R] = \sin(kR)\cos(k_c R) + \cos(kR)\sin(k_c R) \quad (\text{B3})$$

to write

$$C_{\Delta f}(kR) = \frac{1}{\Delta k} \int_{-\Delta k/2}^{\Delta k/2} \left\{ \frac{\sin(kR)\cos(k_c R)}{(k + k_c)R} + \frac{\cos(kR)\sin(k_c R)}{(k + k_c)R} \right\} dk. \quad (\text{B4})$$

Since k varies in the range of Δk , the denominator of the two above integrals can be Taylor expanded as power of k/k_c to give:

$$\frac{1}{kR + k_c R} = \frac{1}{k_c R} \left[1 - \frac{k}{k_c} + \left(\frac{k}{k_c} \right)^2 - \dots \right]. \quad (\text{B5})$$

Using this expansion in (B4) and neglecting high $\Delta k/k$ orders, it can be shown after few manipulations that the correlation function is given by

$$C_{\Delta f}(kR) \sim \text{sinc}\left(\frac{\Delta k R}{2}\right) \left[\frac{\sin(kR)}{kR} - 2 \frac{\cos(kR)}{(kR)^2} \sin^2\left(\frac{\Delta k R}{4}\right) \right], \quad (\text{B6})$$

where the function $\text{sinc}(x)$ is defined by

$$\text{sinc}(x) = \frac{\sin(x)}{x} \quad (\text{B7})$$

and where one has now $k \equiv k_c = (k_1 + k_2)/2$.

The second term of (B6) decreases as R^3 as R increases while its maximum value is at $R=0$ and is given by

$1/8(\Delta k/k)^2$. Therefore, for typical finite frequency bands, the second terms in (B6) can be neglected so the correlation function reads

$$C_{\Delta f}(kR) \sim \text{sinc}\left(\frac{\Delta k R}{2}\right) \frac{\sin(kR)}{kR}. \quad (\text{B8})$$

Numerical integrations of (B1) have shown that this approximation remains valid for frequency bands as large as octave bands.

- ¹L. L. Beranek, *Noise and Vibration Control* (McGraw-Hill, New York, 1988), 2nd ed.
- ²H. Kuttruff, *Room Acoustics* (Applied Science, 1991), 3rd ed.
- ³A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981).
- ⁴T. J. Schultz, "Diffusion in reverberation rooms," *J. Sound Vib.* **16**, 17–28 (1971).
- ⁵A. Abdou and R. W. Guy, "A review of objective descriptors for sound diffuseness," in *Proceedings Issue: Canadian Acoustics*, 43–44 (1994).
- ⁶K. Bodlund, "A new quantity for comparative measurements concerning the diffusion of stationary sound fields," *J. Sound Vib.* **44**, 191–207 (1976).
- ⁷F. Jacobsen, "The diffuse sound field," The Acoustics Laboratory, Technical University of Denmark, Technical Report 27, 1979 (unpublished).
- ⁸R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, Jr., "Measurements of correlation coefficients in reverberant sound fields," *J. Acoust. Soc. Am.* **27**, 1072–1077 (1955).
- ⁹C. G. Balachandran, "Random sound field in reverberant chambers," *J. Acoust. Soc. Am.* **31**, 1319–1321 (1959).
- ¹⁰M. R. Schroeder, "Measurement of sound diffusion in reverberation chambers," *J. Acoust. Soc. Am.* **31**, 1407–1414 (1959).
- ¹¹D. Lubman, "Traversing microphone spectroscopy as a means for assessing diffusion," *J. Acoust. Soc. Am.* **56**, 1302–1304 (1974).
- ¹²M. R. Hodgson, "On the prediction of sound fields in large empty rooms," *J. Acoust. Soc. Am.* **84**, 253–261 (1988).
- ¹³M. R. Hodgson, "On measures to increase room sound-field diffuseness and the applicability of the diffuse-field theory," *J. Acoust. Soc. Am.* **95**, 3651–3653 (1994).
- ¹⁴L. J. Zeng, "The sound distribution in a rectangular reverberation chamber," *J. Acoust. Soc. Am.* **92**, 600–603 (1992).
- ¹⁵M. Koyasu and M. Yamashita, "Evaluation of the degree of diffuseness in reverberation chambers by spatial correlation technique," *Acoust. J. Jpn.* **26**, 132–143 (1971).
- ¹⁶M. Tohyama, A. Suzuki, and S. Yoshikawa, "Correlation coefficients in a rectangular reverberation room," *Acustica* **39**, 51–53 (1977).
- ¹⁷M. Tohyama, A. Suzuki, and S. Yoshikawa, "Correlation coefficients in a rectangular reverberation room experimental results," *Acustica* **42**, 184–186 (1979).
- ¹⁸C. T. Morrow, "Point-to-point correlation of sound pressure in reverberation chamber," *J. Sound Vib.* **16**, 29–42 (1971).
- ¹⁹W. K. Blake and R. V. Waterhouse, "The use of cross-spectral density measurements in partially reverberant sound fields," *J. Sound Vib.* **54**, 589–599 (1977).
- ²⁰C. F. Chien and W. W. Soroka, "Spatial cross-correlation of acoustic pressures in steady and decaying reverberant sound fields," *J. Sound Vib.* **48**, 235–242 (1976).
- ²¹W. T. Chu, "Spatial cross-correlation of reverberant sound fields," *J. Sound Vib.* **62**, 309–311 (1979).
- ²²W. T. Chu, "Comments on the coherent and incoherent nature of a reverberant sound field," *J. Acoust. Soc. Am.* **69**, 1710–1715 (1981).
- ²³R. V. Waterhouse, "Statistical properties of reverberant sound fields," *J. Acoust. Soc. Am.* **43**, 1436–1443 (1968).
- ²⁴W. T. Chu, "Eigenmode analysis of the interference patterns in reverberant sound fields," *J. Acoust. Soc. Am.* **68**, 184–190 (1980).
- ²⁵Y. Kubota and E. H. Dowell, "Asymptotic modal analysis for sound fields of a reverberant chamber," *J. Acoust. Soc. Am.* **92**, 1106–1112 (1992).
- ²⁶"Détermination des niveaux de puissance acoustique émis par les sources de bruit—méthodes de laboratoire en salles réverbérantes pour les sources émettant des fréquences discrètes et des bruits à bandes étroites," ISO standard 3742-1975.
- ²⁷M. Bruneau, *Introduction aux théories de l'acoustique* (Université du Maine Éditeur, Le Mans, France, 1988).
- ²⁸H. Nélisse, O. Beslin, and J. Nicolas, "Panel dynamic response to a reverberant acoustic field," *Am. Inst. Aeron. Astron. J.* **33**, 1590–1596 (1995).
- ²⁹M. R. Schroeder, "The Schroeder frequency revisited," *J. Acoust. Soc. Am.* **99**, 3240–3241 (1996).

Matched-field localization for multiple sources in an uncertain environment, with application to Arctic ambient noise

Michael V. Greening^{a)}

Datavision Computing Services Ltd., 203-1545 Pandora Avenue, Victoria, British Columbia V8R 6R1, Canada

Pierre Zakarauskas^{b)} and Stan E. Dosso^{c)}

Defence Research Establishment Atlantic, Esquimalt Defence Research Detachment, FMO, Victoria, British Columbia V0S 1B0, Canada

(Received 8 June 1995; accepted for publication 30 August 1996)

This paper describes a general method to localize multiple broadband sources in a range- and azimuth-dependent environment when properties of the environment are not well known. Both the source locations and the unknown environmental parameters are included as variables in a simulated annealing inversion that searches for the best match between the measured and modeled fields. The example used to illustrate the method involves localizing an unknown number of ice-ridge building events in an Arctic environment where the bathymetry is poorly known. A number of aspects of this problem are examined in a synthetic study, including the effect of errors in the bathymetry on localization, the importance of simultaneous localization, and determining the number of sources present. The inversion accurately localizes multiple synthetic sources, particularly in cases where the sources are spatially separated and the spectral shape is known. The inversion is also applied to a set of ambient noise measurements recorded in the Lincoln Sea. The results indicate that a small number of ice-ridging sources is sufficient to accurately model the measured fields. Estimates of the number of sources and the source bearings are unambiguous and are reasonably consistent between samples; however, source ranges could not be estimated unambiguously. [S0001-4966(97)00305-6]

PACS numbers: 43.60.Gk, 43.30.Wi, 43.30.Pc, 43.60.Pt [JLK]

INTRODUCTION

This paper describes a general method of applying matched-field processing (MFP) to localize multiple broadband sources when properties of the environment are not well-known. Matched-field processing consists of modeling the acoustic field produced by a simulated source on an array of receivers and “matching” this field to that measured at the array. If the array geometry and environmental properties are known accurately, then the best match will occur when the simulated source is at the same location as the true source. In practice, however, sufficiently complete and accurate measurements of the environment may not be available; this leads to a condition known as mismatch, and can preclude reliable localization using standard MFP methods. In particular, uncertainty in bathymetry can strongly affect localization performance in shallow-water environments.^{1,2} This is a practical problem in the Arctic, where the bathymetry is often poorly known and difficult to measure due to ice cover. A general approach to the problem of environmental mismatch in MFP was developed by Collins and Kuperman.³ This approach, known as focalization, is based on generalizing the localization problem to include uncertain environ-

mental parameters as well as source location in the search for an optimal match with the acoustic data. Collins and Kuperman³ applied this method to correct for mismatch due to uncertainty in the sound-speed profile. Dosso⁴ applied the method of focalization to the problem of source localization in an environment with an incomplete knowledge of bathymetry. He divided the bathymetry into a large number of linear segments on a fixed range partition and used the method of simulated annealing to adjust the depth of the segment end points within assumed bounds.

Localizing multiple sources represents a difficult problem in MFP. When the Bartlett and other low-resolution processors are used, weak sources may be lost in the sidelobes of stronger sources, and cannot be identified when the data are compared to the modeled field for a single source. High-resolution techniques such as maximum-likelihood or eigenvector processing suppress sidelobes but are very sensitive to mismatch in the environmental or array parameters.

Ozard⁵ attempted to resolve multiple sources using MFP on both measured and simulated data. Acoustic data were recorded on a four-element array for a submerged source with a signal-to-noise ratio (SNR)=20 dB towed below a surface source of SNR=0 dB. An eigenvector technique was used and detailed knowledge of the environmental and array properties was available. The bearing of the deep source was determined, with some ambiguity and no estimate of range. In a simulation study, the source bearing and depth could be determined when one source was present; however, source depths were ambiguous when two sources were present.

^{a)}Current address: Cooperative Research Centre for Robust and Adaptive Systems, TSAN 64, P.O. Box 1500, Salisbury, S. A., 5108, Australia.

^{b)}Current address: Department of Ophthalmology, University of British Columbia, 2550 Willow St., Vancouver, BC V5Z 3N9, Canada.

^{c)}Current address: School of Earth and Ocean Sciences, University of Victoria, Victoria, BC V8W 3P6, Canada.

Simulations for a 15-element array and two sources indicated that both source ranges could be resolved, but source depths were ambiguous.

Mirkin and Sibul⁶ used a maximum-likelihood technique to detect and localize two simulated sources using matched-mode processing. They assumed all array and environmental parameters were known exactly and searched for equal-strength sources in both uncorrelated white noise and correlated surface noise. They employed a multistep process in which they first searched for only one source. This source was then nulled and the second source was sought. The second source was then nulled from the original cross spectrum and the first source sought again. This process was repeated until the source locations did not change. With the correlated surface noise 4.8 dB above the uncorrelated white noise, both sources were consistently localized for SNR > 12 dB. The probability of detection decreased with SNR until neither source could be localized for SNR < -3 dB.

Collins *et al.*⁷ used a combination of eigenvector and Bartlett processing to localize multiple moving sources using simulated data. They matched the acoustic field from a source with a single-signal eigenvector using a Bartlett processor. For a single source, the acoustic field and the signal eigenvector are equal (within a scale factor) and the technique works well. However, a potential problem exists for multiple sources, since the field produced by an individual source is a linear combination of the signal eigenvectors, and may not match a single-signal eigenvector. Collins *et al.*⁷ showed that multiple sources can be detected and localized provided the signals are of unequal levels, are uncorrelated, and do not remain in each other's sidelobes (e.g., moving sources). None of the studies cited here examined environmental mismatch or the problem of determining the number of sources.

This paper considers the problem of localizing multiple sources in a range- and azimuth-dependent environment in

cases where the bathymetry is not well-known. A broadband MFP inversion algorithm is developed which searches simultaneously for the location (range and bearing) and relative strength of multiple sources, and includes corrections to the bathymetry model along the propagation path to each source in the inversion process. The approach is illustrated with synthetic examples representing a distribution of active ice-ridge building events in an Arctic environment where the bathymetry is poorly known. A number of aspects of the inverse problem are examined in the synthetic study, including the effect of errors in the bathymetry on localization, the importance of simultaneous localization, and the problem of determining the number of sources present. The inversion method is then applied to a set of Arctic ambient noise measurements collected in the Lincoln Sea in an attempt to investigate the spatial distribution of active ice-ridge building events. Section I of this paper describes the matched-field inversion algorithm and its implementation. Section II considers synthetic examples, and Sec. III applies the inversion to Arctic ambient noise measurements.

I. THE INVERSION ALGORITHM

A. The modified Bartlett processor

This section provides a mathematical description and details of the implementation of the MFP inversion used to localize multiple broadband sources in uncertain bathymetry. The goal of MFP is to minimize the mismatch between the measured cross-spectral matrix R and a modeled cross-spectral matrix M . Broadband MFP decreases the number and amplitude of sidelobes, since sidelobes tend to occur at different ranges for different frequencies. For a broadband source, cross-spectral matrices may be computed at individual frequencies and the mismatch or error E defined as⁸

$$E = 1 - \left[\frac{\sum_{p=1}^H \sum_{q=p+1}^H \sum_{f=f_l}^{f_u} |R_{pq}(f)| M_{pq}^*(f)}{(\sum_{p=1}^H \sum_{q=p+1}^H \sum_{f=f_l}^{f_u} |R_{pq}(f)|^2)^{1/2} (\sum_{p=1}^H \sum_{q=p+1}^H \sum_{f=f_l}^{f_u} |M_{pq}(f)|^2)^{1/2}} \right], \quad (1)$$

where H is the number of hydrophones, f_l and f_u are the lower and upper limits of the frequency band, $R_{pq}(f) = \langle R_p(f) R_q^*(f) \rangle$ is the cross spectrum of the measured data for receivers p and q at frequency f , $\langle \cdot \rangle$ denotes the expected value, $R_p(f)$ is the Fourier transform of the acoustic field received at hydrophone p , $M_{pq}(f) = M_p(f) M_q^*(f)$ is the modeled cross spectrum for receivers p and q , and $M_p(f)$ is the Fourier transform of the modeled acoustic field at hydrophone p . Equation (1) is a variation of the Bartlett processor adapted to include multiple frequencies and to delete the autospectra of the cross-spectral matrices. Tests have shown that summing coherently over both frequency and hydrophones and deleting the autospectra results in smaller Bartlett errors and lower ambiguity than summing

incoherently and/or including the autospectra.^{8,9} The Bartlett processor was chosen because it is less sensitive to mismatch in the environment than high-resolution processors such as maximum-likelihood or eigenvector techniques, and because it is easily adapted to include multiple broadband sources (other robust processing techniques, such as the inverse beamforming processor,¹⁰ could also be adapted to this problem). High-resolution methods are often used for multiple sources to reduce the overlap between the mainlobe of one source and sidelobes of other sources; however, these techniques tend to be very sensitive to array or environmental mismatch.

The modeled acoustic field for multiple sources may be found by (coherently) adding the sources as:

$$M_q(f) = \sum_{i=1}^S a_i(f) p(r_i, z_i, \theta_i, z_q, f), \quad (2)$$

where $a_i(f)$ is the source pressure level of source i at frequency f , S is the number of sources, and $p(r_i, z_i, \theta_i, z_q, f)$ is the complex pressure received at receiver depth z_q due to a unit strength source located at range r_i , depth z_i , and bearing θ_i . The modeled acoustic field may also be determined by specifying pressure levels at the receiver, instead of the source, as:

$$M_q(f) = \sum_{i=1}^S \left[\frac{b_i(f) p(r_i, z_i, \theta_i, z_g, f)}{[\sum_{j=1}^H |p(r_i, z_i, \theta_i, z_j, f)|^2]^{1/2}} \right], \quad (3)$$

where $b_i^2(f)$ is the total power of source i received at all hydrophones. Equation (3) compares the relative received levels at the array and was used for all inversions in this paper. In some cases, the true source levels may be of interest and Eq. (2) can be used. Note, however, that using Eq. (2) causes the source level and source range parameters to be coupled (both affect the received level) and the inversion becomes much more difficult. An alternative approach for obtaining source levels would be to use Eq. (3) and add the propagation loss after the sources are localized.

The acoustic pressure at receiver depth z_q due to a source at (r_i, z_i, θ_i) was calculated using an $N \times 2D$ adiabatic normal mode model:

$$p(r_i, z_i, \theta_i, z_q) = \sum_{j=1}^N \phi_j[d(r_i, \theta_i), z_i] \phi_j[d(0, 0), z_q] \frac{\exp(i \int_0^{r_i} k_j[d(r, \theta_i)] dr)}{\sqrt{\int_0^{r_i} k_j[d(r, \theta_i)] dr}}, \quad (4)$$

where N is the number of modes that propagate from the source to the receiver, $\phi_j[d(r, \theta), z]$ is the j th mode function at depth z for a water depth of $d(r, \theta)$, and $k_j[d(r, \theta)]$ is the j th wave number. The mode functions ϕ_j and wave numbers k_j were calculated for 20-m increments in water depth using the KRAKENC normal mode program developed by Porter.¹¹ These values were then stored in look-up tables for fast reference, and linear interpolation was used to determine values for intermediate water depths. Use of the adiabatic normal mode model with modal lookup and interpolation provides highly efficient repetitive computations of the acoustic matching fields.⁴ This is of key importance to the multidimensional inversion developed here, which requires the matching fields be modeled an extremely large numbers of times (inversions presented in Sec. III required $\sim 10^6$ calculations of the matching fields). The adiabatic normal mode model is applicable to environments with moderate range dependence where the effects of mode coupling are small, and should be appropriate for localization in uncertain bathymetry provided the bottom slopes remain sufficiently small. The use of a coupled-mode propagation model would alleviate this restriction. However, coupled-mode models are computationally intensive and are not amenable to the use of look-up tables for efficiency, and would therefore result in an impractical inversion algorithm.

B. Simulated annealing inversion

The algorithm developed here for matched-field localization of multiple sources with uncertain bathymetry searches simultaneously for the range, azimuth, and relative strength of a specified number of sources, together with corrections to the bathymetry model along the propagation path to each source. This requires that an extremely large parameter space be examined to find the parameter set which minimizes the Bartlett error between the measured and model fields [Eq. (1)]. The parameter search is carried out using the method of simulated annealing which provides an efficient stochastic search algorithm that inherently avoids becoming trapped in local suboptimal solutions.^{3,4,12} Simulated annealing involves a series of iterations in which the unknown parameters are perturbed a number of times. After each perturbation, the change in mismatch ΔE is calculated. If the mismatch has decreased ($\Delta E < 0$), the new parameter configuration is accepted. The case of an increase in mismatch ($\Delta E > 0$) is treated probabilistically, with the new configuration accepted with a probability P drawn from the Boltzmann distribution

$$P(\Delta E) = \exp(-\Delta E/T), \quad (5)$$

where T is a controlling parameter analogous to temperature in the physical process of annealing.¹² Accepting some perturbations which increase E allows the algorithm to escape from local minima in the search space. Decreasing T with successive iterations decreases the probability of accepting an increase in mismatch, and the algorithm eventually converges to a solution which should approximate the global minimum.

Two factors involved in developing an efficient and effective simulated annealing algorithm are the method of decreasing the temperature T , known as the annealing schedule, and the method of perturbing the unknown parameters. The annealing schedule used here starts with an initial temperature T_0 that allows at least 90% of all perturbations to be accepted. A number of perturbations η are then performed before decreasing the temperature according to $T_{j+1} = \alpha T_j$, where $\alpha < 1$. The process is terminated when further temperature steps do not result in a lower Bartlett error. The values η and α are estimated depending on the difficulty of the inversion. This tends to be problem-specific and requires experimentation with different values. If both the measured and modeled cross-spectral matrices R and M are generated synthetically, then η and α can be chosen large enough that a Bartlett error of essentially zero will occur. However, this may result in a prohibitively large number of iterations and correspondingly large computation times, and thus smaller values of η and α may be used which result in a finite, but small, Bartlett error and acceptably accurate results for the source and environmental parameters. With measured data where the true value of the parameters are unknown, η and α can be chosen so that further increases in their values do not decrease the final Bartlett error E .

The method of perturbing the parameters in the simulated annealing algorithm can have a major effect on the efficiency of the inversion. After experimenting with a number of perturbation schemes, we have found that the follow-

ing general approach works well. With the exception of the bathymetry, the parameters were perturbed one at a time. This allowed the algorithm to converge for a sensitive parameter while continuing to search for less sensitive parameters. For bathymetry, either one or two water depths were perturbed at a time. Perturbing two water depths allowed the possibility of changing the bathymetry while still maintaining approximately the same wave-number integrals [Eq. (4)], and was found to result in significantly smaller Bartlett errors for the same number of iterations. For each perturbation, a single source was randomly selected and an equal probability given to perturbing either the bearing, range, relative source strength, or bathymetry in the direction of the source. The parameter chosen was then changed in one of two ways. Either a new random value was chosen from the entire allowed range for that parameter, or the parameter was perturbed so as to remain relatively close to the previous value according to

$$f_{j+1} = f_j + \epsilon^3 \Delta, \quad (6)$$

where f_{j+1} is the new (perturbed) parameter, ϵ is a random number drawn from a uniform distribution on $[-1, 1]$, and Δ is the maximum perturbation allowed for the parameter in one step. Cubing ϵ has the effect that small perturbations are more likely, but changes as large as Δ are possible.¹³ As in determining the annealing schedule, determining an appropriate value for Δ for each parameter can require some experimentation. In addition to this general perturbation scheme, other methods of changing the parameters can be used for specific cases. Some of these are described for the synthetic test case in the following section.

II. SYNTHETIC STUDY

A. Arctic test case

A number of aspects of the problem of localizing multiple sources in uncertain bathymetry are investigated in this section using synthetic acoustic data. This synthetic study was designed to simulate ambient noise measurements collected in the Lincoln Sea which are analyzed in the following section. The environment and array geometry of the acoustic experiment were adopted for the synthetic study and are briefly summarized here; the recording and inversion of the ambient noise samples are considered in detail in Sec. III.

The hydrophone array was located in 430 m of water and consisted of vertical and horizontal subarrays of 24 and 7 sensors, respectively, as shown in Fig. 1. The horizontal subarray included one off-axis hydrophone to break the left-right symmetry about the main axis of the array. The bathymetry and ocean sound-speed profile in the study region are shown in Figs. 2 and 3, respectively. The bathymetry were taken from the chart of the Arctic basin published by the National Research Laboratory (NRL).¹⁴ This chart was digitized out to a range of 400 km from the array site along radial paths every 1° in azimuth. Synthetic acoustic data and matching fields were computed for the frequency band 10–30 Hz in 5-Hz increments for this environment and array using the $N \times 2D$ adiabatic normal mode model described in Sec. I A.

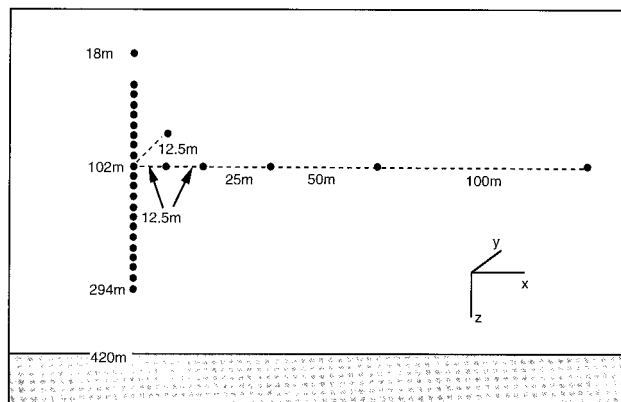


FIG. 1. Array configuration. The vertical sub-array consisted of 24 hydrophones at 12-m separations from 18- to 294-m depth (hydrophones at 30 and 42 m did not operate). The horizontal sub-array consisted of seven hydrophones, separated as shown, at 102-m depth with one off-axis hydrophone.

The applicability of the adiabatic approximation to an environment very similar to that considered here was investigated in a simulation study by Zala and Ozard.¹⁵ They compared an adiabatic normal mode model to a parabolic equation (PE) model which included three Padé approximation terms and is considered an accurate range-dependent propagation model.¹⁶ The environment they considered involved an Arctic sound-speed profile and a realistic geoacoustic model with a bottom that sloped downward from a vertical array in 500 m of water. They compared the two propagation models as a function of bottom slope at a frequency of 20 Hz, and found excellent agreement for slopes $< 3^\circ$. The largest bottom slopes along radial paths out from the array site in the digitized bathymetry of Fig. 2 are $< 2.4^\circ$, with slopes $< 1^\circ$ within 20 km of the array site. Hence, the adiabatic normal mode model should be applicable to this environment.

In the simulated annealing inversions, the parameter bounds and maximum perturbations Δ adopted were as fol-

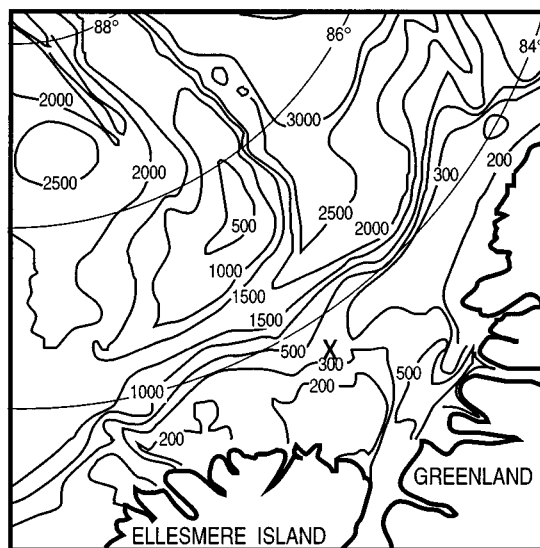


FIG. 2. Bathymetry of the Lincoln Sea and Arctic basin. Location of the array site is marked by the "X."

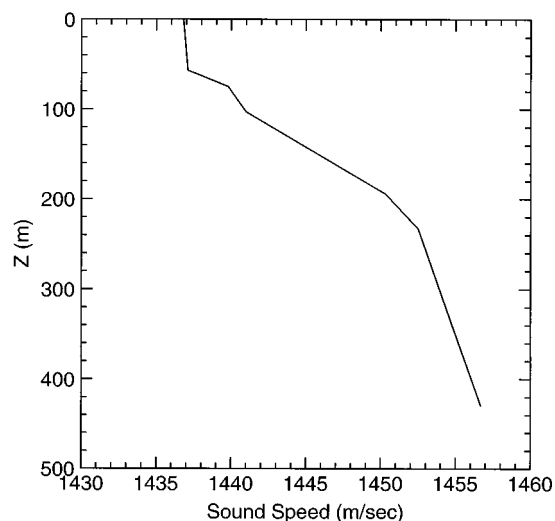


FIG. 3. Sound-speed profile in the water column at the array site. Sound speed for deeper water is assumed to increase linearly with depth to 1520 m/s at 3000-m depth.

lows. For source bearing, a value of $\Delta = 3^\circ$ was used and bearings were allowed to vary over 0° – 360° . For source ranges, a value of $\Delta = 5$ km was used and source ranges were allowed to vary over 0–400 km. (Since active pressure ridges are irregular line sources which can be 5 km or more in length,¹⁷ this value for Δ is appropriate to keep small range perturbations on the same scale as the source dimensions.) For relative source pressure levels, $\Delta = 3$ dB was used and received levels were varied within ± 30 dB of an arbitrary reference. For bathymetry, the segment end points were changed in depth with $\Delta = 5$ m and a maximum change of ± 50 m from starting depths taken from the NRL chart.¹⁴

In addition to perturbing the parameters according to Eq. (6), including several other types of perturbations was found to provide significantly improved efficiency. By examining ambiguity surfaces, it was found that large sidelobes exist in the Bartlett error at 10–20-km intervals in range. Therefore, an option was included of changing the source range by a random perturbation of 5–25 km to include large jumps which correspond to transitions between sidelobes. Because the horizontal array included only one off-axis hydrophone (Fig. 1), a strong ambiguity exists between a given source bearing and its reflection about the main axis of the array. Therefore, an option was included which consisted of reflecting the source bearing about the array axis.

Finally, since the bathymetry model is a function of source bearing, when a new bearing is accepted in the simulated annealing algorithm, the current bathymetry may not lie within the assumed bounds for that bearing. In this case, a new bathymetry model would need to be determined with no carry-over information from the previous iterations. Due to the sensitivity of the mismatch to bathymetry, this can result in few changes in bearing being accepted. To overcome this difficulty, when a new source bearing is accepted, the bathymetry is not changed for that iteration. Future iterations then select new random values for the bathymetry from within the bounds for that bearing until a match is accepted. This approach has proved to be effective, and we have not

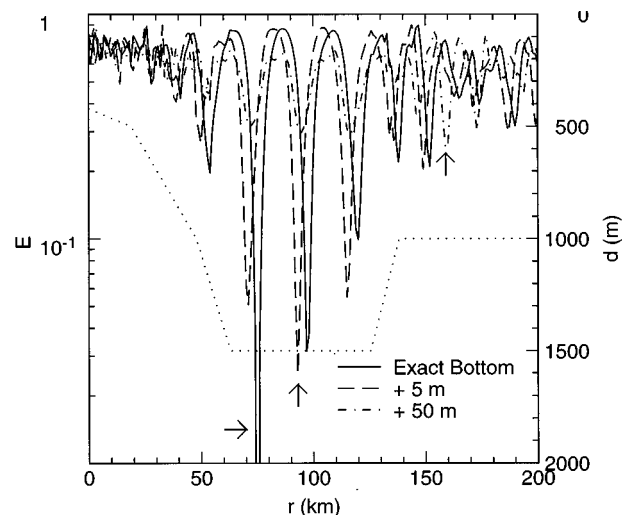


FIG. 4. Bartlett error E for a simulated source at 75-km range. Solid line shows E when the exact bathymetry (dotted line) is known and indicates a perfect match at 75 km. Dashed line shows E when water depths greater than 1000 m are overestimated by 5 m. Dash-dot line shows E when depths greater than 1000 m are overestimated by 50 m. Arrows indicate minimum error for each curve.

found any cases where the final bathymetry did not lie within the assumed bounds.

B. Effects of bathymetry

Uncertainty in bathymetry represents a major source of mismatch when applying MFP in Arctic environments. Figure 4 shows the Bartlett error E as a function of range for a source at 75-km range and -20° bearing (bearings given clockwise from north). If the true bathymetry (dotted line) is used in computing the matching fields, a perfect match (zero mismatch) occurs at 75-km range with sidelobes at 54, 97, 120, 138, and 152 km. If the matching fields are computed using all contour lines at ≥ 1000 -m depth offset by 5 m, the smallest mismatch is 0.025 and occurs at 93 km. If the same contour lines are offset by 50 m, the best mismatch is 0.25 and occurs at 159 km. Thus, errors in water depth can result in large errors in source range estimation.

Because the bathymetry in the Arctic is often poorly known, standard MFP methods may fail. To overcome this problem, the bathymetry along the propagation path can be included as a variable to be determined in the inversion algorithm. Equation (4) indicates that, within the adiabatic approximation, the acoustic pressure depends on the water depth at the source range through the mode functions, and on the bathymetry over the propagation path by the range integrals of the wave numbers. Thus, the true bathymetry is not actually required to reproduce the measured acoustic fields. All that is required is an acoustically equivalent bathymetry model which has the correct water depth at the source and results in the correct wave-number integrals.⁴ Zakarauskas *et al.*¹⁸ have shown that this equivalent bathymetry model can be represented adequately by a small number of linear segments. The method used here to represent bathymetry is to define a set of segments with one end point corresponding

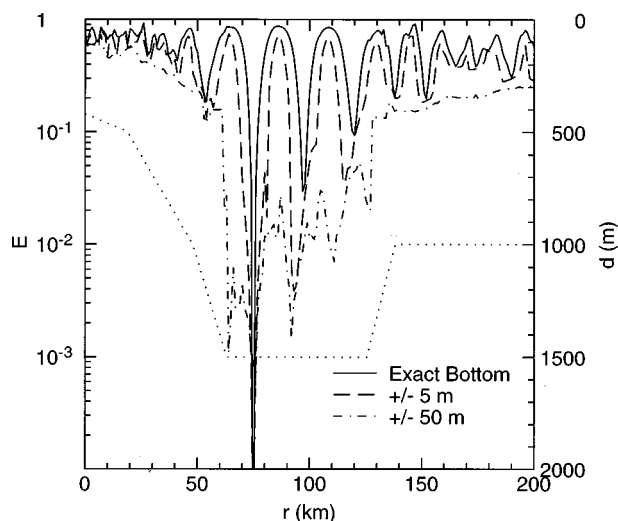


FIG. 5. Bartlett error E for a simulated source at 75-km range. Solid line shows E when the exact bathymetry (dotted line) is known. Dashed line shows E when an uncertainty of ± 5 m from the true bathymetry is allowed. Dash-dot line shows E when the uncertainty is ± 50 m. The true source range is determined in each case.

to an isobath of the bathymetry chart and the other end point midway between isobaths. The depth of each end point is then allowed to vary within assigned bounds of the chart values.

Figure 5 compares the Bartlett error obtained when the true bathymetry is known to cases where an uncertainty in the bathymetry of ± 5 m and ± 50 m is assumed. In the cases of uncertain bathymetry, the curves represent the best match at each range obtained using the simulated annealing algorithm to search over bathymetry within the assumed uncertainty. Figure 5 shows that in each case the best match occurs at the true source location. The figure also shows that the mainlobe and sidelobes broaden and deepen as the uncertainty in the bathymetry is increased.

In Fig. 5, the true bathymetry is included in the inversion search space in each case. When the true bathymetry is not within the search space, the true source location may or may not be correctly determined. The effects of bathymetry errors at the source location and over the propagation path are investigated in Figs. 6 and 7. These figures were produced using synthetic acoustic fields computed for a source at 75-km range using the (true) bathymetry shown by the dotted line. However, in computing the matching fields for the inversions, the bathymetry was constrained within ± 50 m of initial bathymetry estimates (solid and dashed lines) which contain offsets of up to 100 m from the true bathymetry. In Fig. 6, two initial bathymetry estimates were used which fluctuate about the true bathymetry (sometimes shallower, sometimes deeper). The water depth at the source range (75 km) for the solid-line bathymetry estimate is within 50 m of the true value, and hence the true value is included in the search space. The corresponding Bartlett error curve has a well-defined minimum at the true source range of 75 km. The water depth at the source range for the dashed-line bathymetry estimate differs from the true value by more than 50 m, and hence the true value is not included in the search space. For this case, the minimum of the Bar-

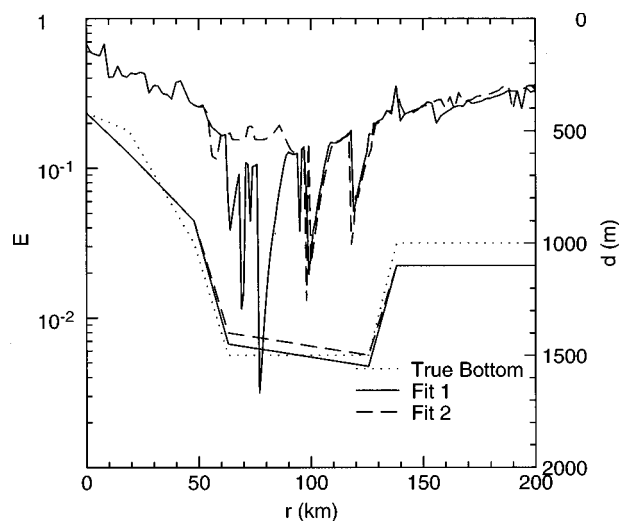


FIG. 6. Bartlett error E for a simulated source at 75-km range when the search space of the bathymetry does not include true bathymetry (dotted line). Water depth errors of up to 100 m are included (sometimes shallower, sometimes deeper), but an uncertainty of only ± 50 m is allowed in the inversion. For fit 1 (solid line), the true water depth at the source range is within the search space; for fit 2 (dashed line), the true depth at the source range is not within the search space.

tle error curve is significantly higher and is located at an incorrect range of 98 km.

In Fig. 7, the initial estimates of the bathymetry are consistently deeper than the true bathymetry. As in Fig. 6, the solid-line bathymetry estimate is within 50 m of the true bathymetry at the source range, while the dashed-line estimate is not. For both cases in Fig. 7, the correct source location is not determined, and even source detection is uncertain.

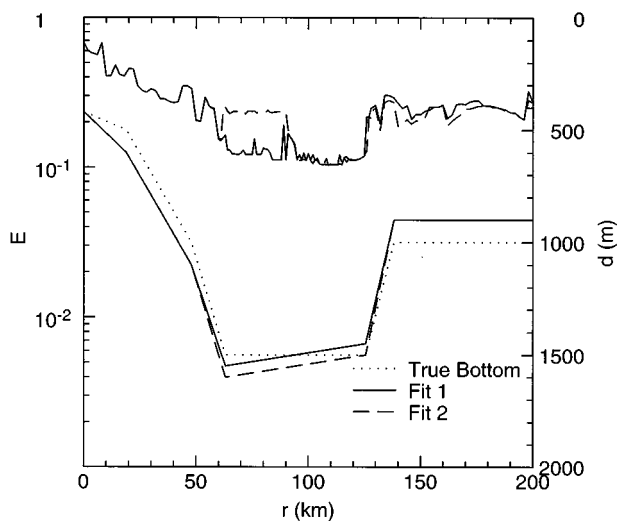


FIG. 7. Bartlett error E for a simulated source at 75-km range when the search space of the bathymetry does not include the true bathymetry (dotted line). Water depth errors of up to 100 m are included (depths are always overestimated), but an uncertainty of only ± 50 m is allowed in the inversion. For fit 1 (solid line), the true water depth at the source range is within the search space; for fit 2 (dashed line), the true depth at the source range is not within the search space.

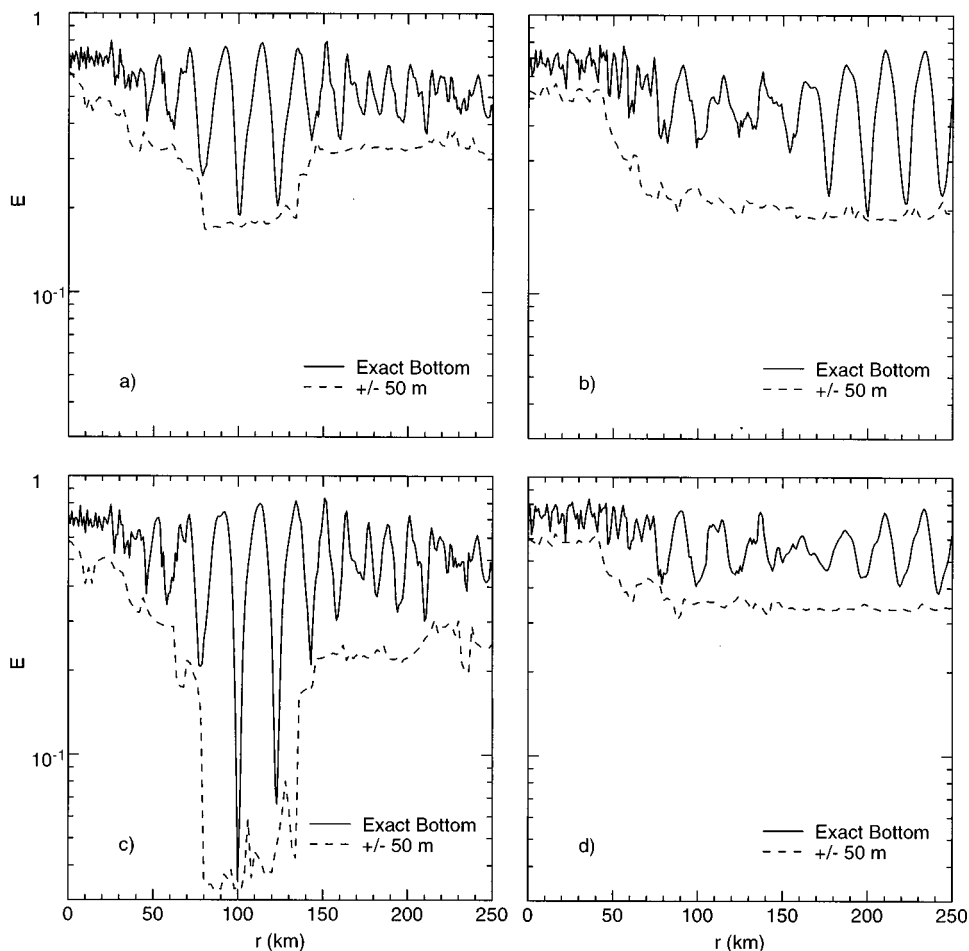


FIG. 8. Bartlett error E when two sources are present, but matching fields are computed for a single source. The solid line represents E when the exact bathymetry is used for the matching fields; the dotted line represents E when the bathymetry is uncertain by ± 50 m and is included in the inversion. Sources are located at -35° , 100 km and $+35^\circ$, 200 km (the bathymetry is taken from Fig. 2). In (a) and (b) the sources are of equal received level; in (c) and (d) the received level of the source at $+35^\circ$ is 5 dB below that of the source at -35° . In (a) and (c) a single source is sought at -35° , while in (b) and (d) a source is sought at $+35^\circ$.

C. Simultaneous inversion for multiple sources

Detecting and localizing multiple sources is a difficult problem in MFP. Previous studies have applied high-resolution techniques with good or exact knowledge of the environment.^{5–7} We have instead chosen to use a variation of the low-resolution Bartlett processor [Eq. (1)] because it is less sensitive to environmental mismatch, such as uncertain bathymetry, and because it is easily adapted to include multiple broadband sources. In standard Bartlett processing, the measured cross spectrum is compared to the modeled field produced by a single source at some trial location. The trial locations which produce the highest matches are assumed to correspond to source locations. With exact bathymetry and sources of equal received levels that are well-separated spatially, the locations of two sources can be determined in this manner, as shown in Fig. 8(a) and (b) by the solid lines. However, with an uncertainty in the bathymetry of ± 50 m, source localization is poor [dashed lines in Fig. 8(a) and (b)] because of the broadening of the sidelobes. As in Figs. 5–7, the curves for uncertain bathymetry represent the smallest Bartlett error found for that range when the bathymetry is allowed to vary within its search interval. If the received

level of one signal is lower than the other, the weaker signal can become obscured by sidelobes of the stronger source, as shown in Fig. 8(c) and (d).

A technique used to localize multiple sources in plane-wave analysis is to search for a single source, remove the effect of the source from the data, then continue searching for the next source.¹⁹ A similar technique, known as prewhitening, is also used to remove the effects of long-term stationary sources from acoustic data.²⁰ An advantage of searching for only one source at a time is that the search space is greatly reduced compared to searching simultaneously for all sources. For example, consider searching for the range, bearing, and relative source level of three sources, together with corrections to a bathymetry model consisting of four segments for each source. Searching for one source at a time requires three passes on a seven-dimensional space (source level, range, bearing, and end-point depths of each bathymetry segment). Searching for the three sources simultaneously requires searching a 21-dimensional space. Although searching for one source at a time greatly reduces the search space and computation time, it does not always result in accurate source localization. This is illustrated in Table I,

TABLE I. Bartlett error E and mismatch M in source parameters for two synthetic sources when localizing sources one at a time. The two input sources are at ranges and bearings of $(R_1, B_1) = (100 \text{ km}, -35^\circ)$ and $(R_2, B_2) = (200 \text{ km}, -20^\circ)$. Results are given for two different initializations of the inversion algorithm and different received levels L_2 for the second source (level of the first source is 0 dB). An uncertainty of $\pm 50 \text{ m}$ was used for bathymetry.

| L_2 (dB) | MR_1 (km) | MB_1 (deg) | MR_2 (km) | MB_2 (deg) | ML_2 (dB) | E |
|---------------|----------------|-----------------|----------------|-----------------|----------------|---------|
| 0.0 | 14.0 | -4.1 | -1.6 | 1.1 | -1.04 | 0.20E-1 |
| 0.0 | -0.5 | -3.4 | 5.1 | -2.2 | 0.5 | 0.22E-1 |
| -3.0 | 30.7 | -2.0 | -0.4 | 1.7 | -1.6 | 0.18E-1 |
| -3.0 | 14.2 | -2.5 | -2.3 | 1.6 | -1.2 | 0.10E-1 |
| -5.0 | 15.6 | -1.4 | -1.3 | 1.2 | -0.3 | 0.56E-2 |
| -5.0 | 31.6 | -1.2 | -46.7 | 1.6 | -0.5 | 0.13E-1 |
| -10.0 | 36.1 | -0.9 | -89.1 | -5.8 | -0.3 | 0.76E-2 |
| -10.0 | 16.4 | -0.2 | -55.2 | 0.3 | -1.4 | 0.91E-3 |

which summarizes the results of a number of inversions for two sources in uncertain bathymetry when the sources are localized one at a time. The difficulty with localizing in this manner is that the measured field is produced by multiple sources, but is compared to matching fields produced by a single source on each pass. The sidelobes of the multiple sources may combine to produce the best match for a single source at a location where no source exists (this is a well-known limitation in plane-wave beamforming). When both sources are localized simultaneously, the estimated source locations and levels are accurate and the resulting mismatch errors are small, as shown in Table II. All of the inversions summarized in Tables I and II were performed using the same annealing schedule.

D. Localization of an unknown number of sources

Two obvious problems in searching for multiple sources involve estimating the number of sources which contribute significantly to the total acoustic field, and investigating how the localization results are affected when an incorrect number of sources is sought. A straightforward method to estimate the number of sources is to examine the minimum Bartlett error E returned by the inversion algorithm as a function of the number of sources sought in the inversion. Figure 9

TABLE II. Bartlett error E and mismatch M in source parameters for two synthetic sources when localizing both sources simultaneously. The two input sources are at ranges and bearings of $(R_1, B_1) = (100 \text{ km}, -35^\circ)$ and $(R_2, B_2) = (200 \text{ km}, -20^\circ)$. Results are given for two different initializations of the inversion algorithm and different received levels L_2 for the second source (level of the first source is 0 dB). An uncertainty of $\pm 50 \text{ m}$ was used for bathymetry.

| L_2 (dB) | MR_1 (km) | MB_1 (deg) | MR_2 (km) | MB_2 (deg) | ML_2 (dB) | E |
|---------------|----------------|-----------------|----------------|-----------------|----------------|---------|
| 0.0 | 0.4 | -0.5 | 0.6 | 0.4 | 0.1 | 0.15E-3 |
| 0.0 | 1.3 | -0.3 | 1.0 | 0.3 | 0.0 | 0.29E-3 |
| -3.0 | -17.3 | -0.5 | 2.1 | 0.0 | 0.0 | 0.41E-3 |
| -3.0 | -0.4 | 0.3 | 0.2 | 0.3 | 0.0 | 0.14E-3 |
| -5.0 | 0.2 | -0.1 | 1.5 | -0.3 | 0.2 | 0.18E-3 |
| -5.0 | 0.1 | 0.7 | 1.1 | 0.0 | -0.2 | 0.28E-3 |
| -10.0 | 0.2 | -0.3 | 2.3 | 0.7 | 0.8 | 0.13E-3 |
| -10.0 | 0.9 | 0.4 | 7.2 | 0.9 | 0.7 | 0.26E-3 |

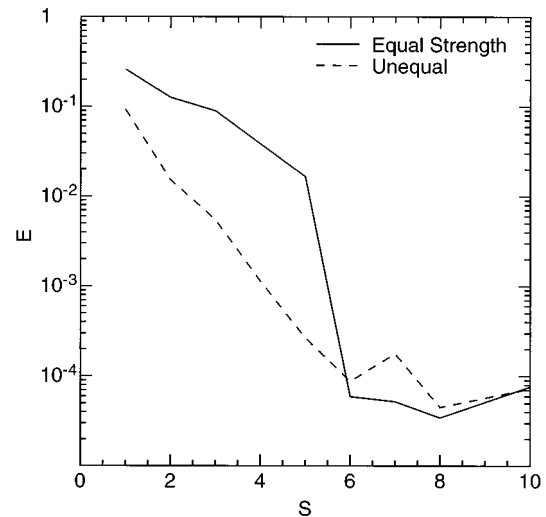


FIG. 9. Bartlett error E as a function of the number of sources S sought in the inversion when six sources of equal (solid line) or unequal (dashed line) received level are used as input. An uncertainty of $\pm 50 \text{ m}$ was used for bathymetry.

shows the Bartlett error as a function of the number of sources sought in the inversion when six sources of equal (solid line) or unequal (dashed line) received level were used as input to compute synthetic acoustic data [the source distributions are given in Figs. 10(a) and 11(a), respectively]. For both cases in Fig. 9, the Bartlett error decreases as the number of sources sought increases to six. Increasing the number of sources beyond six does not significantly decrease the Bartlett error.

Figure 10 shows the estimated source locations and relative levels (circle size) determined by the inversion algorithm when six sources of unequal received level are used as input and 1–10 sources are sought. In this example, a flat source spectrum was used, and the spectral shape was assumed to be known. The search interval for the bathymetry in the inversion was $\pm 50 \text{ m}$. When the correct number of sources is sought [Fig. 10(d)], excellent estimates of the source locations and relative received levels are obtained. When the number of sources is overestimated [Fig. 10(e) and (f)], excellent estimates of the input source locations and levels are still obtained, and additional weak sources are included to account for the required total number of sources. These additional sources do not contribute significantly to the total acoustic field. When the number of sources is underestimated, the stronger sources are correctly localized, as shown in Fig. 10(a)–(c).

Figure 11 shows an example similar to Fig. 10, except that the six sources are of equal received level. In Fig. 11, excellent results are obtained when the number of sources sought in the inversion is equal to or exceeds the number of input sources. However, if the number of sources is underestimated, incorrect source locations can occur, as shown in Fig. 11(a)–(c). The Bartlett errors for Figs. 10 and 11 are shown in Fig. 9, which indicates errors of $E \approx 0.0001$ when the correct number of sources (or more) are included.

Figures 9–11 illustrate detection and localization of multiple sources which are spatially separated and have

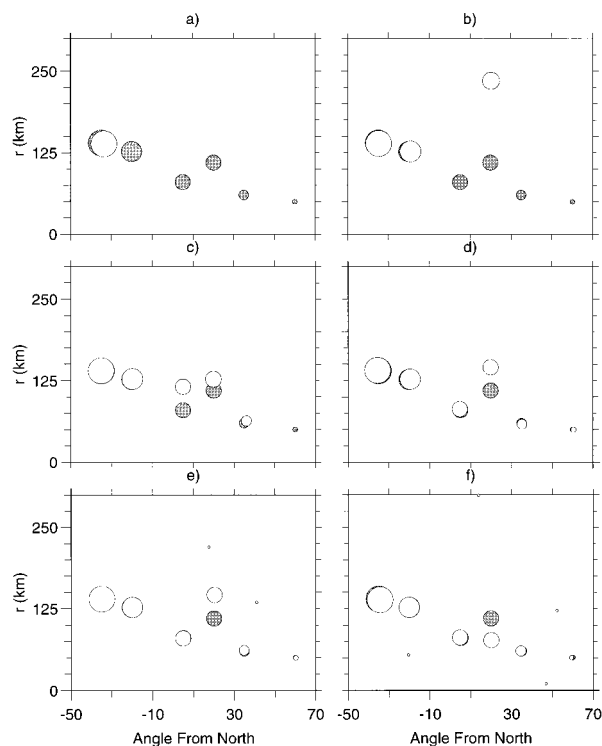


FIG. 10. Results of searching for multiple sources. Six sources of received level 0, -4, -8, -8, -12, and -16 dB are used as input (shaded circles). The results of searching for 1, 3, 5, 6, 8, and 10 sources (open circles) are shown in (a)–(f); respectively. Circle size indicates relative received levels. Bartlett errors E are shown in Fig. 9. An uncertainty of ± 50 m was used for bathymetry.

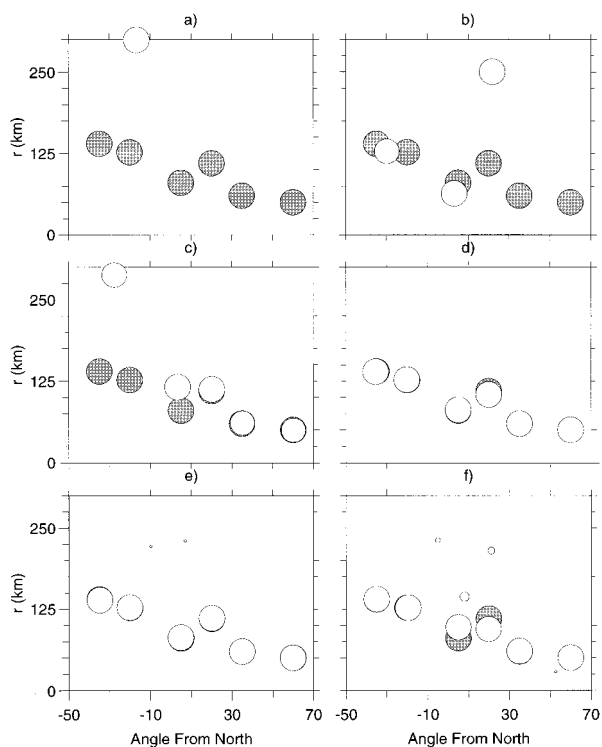


FIG. 11. Results of searching for multiple sources. Six sources of equal received level are used as input (shaded circles). The results of searching for 1, 3, 5, 6, 8, and 10 sources (open circles) are shown in (a)–(f), respectively. Circle size indicates relative received levels. Bartlett errors are shown in Fig. 9. An uncertainty of ± 50 m was used for bathymetry.

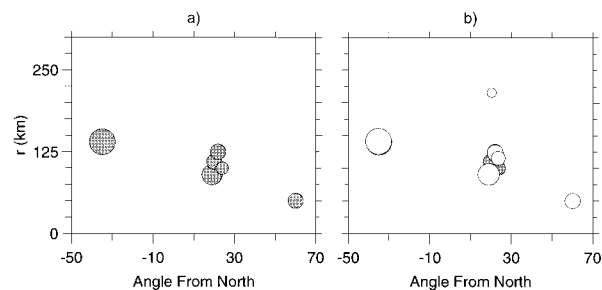


FIG. 12. Results of searching for sources at small azimuthal separation. Input sources are shown in (a) and inversion results in (b). The Bartlett error is $E=0.0002$. An uncertainty of ± 50 m was used for bathymetry.

known spectral shapes. If the spectral shape is unknown or a number of sources are closely spaced in azimuth, the localization is less successful, as shown in Figs. 12 and 13. When four closely spaced sources occur [Fig. 12(a)], the bearings of the sources are all correctly determined, but the ranges of two sources are incorrect, one by approximately 15 km and the other by over 100 km [Fig. 12(b)]. The final Bartlett error is still small, $E=0.0002$. When the spectral shape cannot be assumed to be known, the number of dimensions in the search space increases by a factor of $S \times (N_F - 1)$ where S is the number of sources and N_F is the number of frequencies used in the broadband Bartlett processor. Figure 13 shows the results of searching for six sources when the spectral shapes are unknown (circle size indicates the average power over the frequency band). The resulting source distribution [Fig. 13(b)] is quite good; however, two of the sources have mismatches in range of 18 and 87 km and the weakest source is not detected. The final Bartlett error for this example is $E=0.001$.

III. INVERSION OF ARCTIC AMBIENT NOISE

A. Experiment and data

In this section, MFP inversion for multiple sources is applied to a set of ambient noise measurements recorded in the Canadian high Arctic. The sources of ambient noise in the Arctic are generally of a different nature than in open water. In ice-covered waters there is usually no shipping noise, wave-wave interaction, or bubble clouds as in open water. However, the ice canopy generates noise at low frequencies by cracking and the creation of pressure ridges. Although the properties of ice cracking sources have been

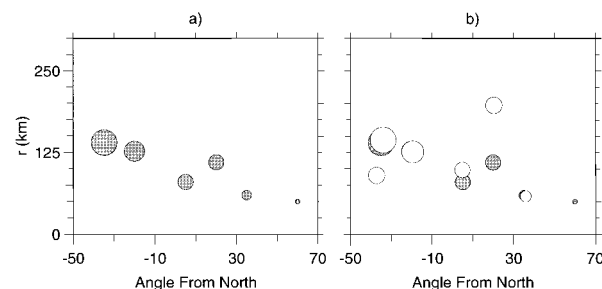


FIG. 13. Results of searching for sources with unknown spectral shapes. Input sources are shown in (a) and inversion results in (b). The Bartlett error is $E=0.001$. An uncertainty of ± 50 m was used for bathymetry.

TABLE III. Time of recording (Coordinated Universal Time) of the six 2-min ambient noise samples. All recordings were made in April 1988.

| Sample | Day | Hour | Minutes |
|--------|-----|------|---------|
| A | 11 | 23 | 23–25 |
| B | 13 | 10 | 03–05 |
| C | 15 | 01 | 41–42 |
| D | 15 | 17 | 48–50 |
| E | 15 | 18 | 47–49 |
| F | 15 | 23 | 07–09 |

analyzed extensively,^{21–28} a number of important questions remain about ice ridging. Pritchard²⁹ found a high correlation between low-frequency ambient noise and the wind and current shearing stresses acting on the ice. These measurements have led to the general belief that low-frequency ambient noise in the Arctic is caused by large-scale ice motion and the resulting pressure ridging. Apart from the recording of nearby ridges of opportunity,^{17,30} little is known of the spatial distribution of pressure ridges which are active at any given time. Greening and Zakarauskas³⁰ showed that measured ambient noise spectra are compatible with a small number of active ridges producing most of the ambient noise power at frequencies below 40 Hz.

The ambient noise measurements analyzed here were recorded using an array of hydrophones with both vertical and horizontal extent (Fig. 1) suspended from the Arctic pack ice. The array was located on the continental shelf in 420 m of water in the Lincoln Sea off the north coast of Ellesmere Island (Fig. 2). Full details of the environment at the experiment site are available in previous papers^{28,30} and are summarized here. The ice thickness varied from approximately 2 m on refrozen leads to ~7 m on multiyear ice floes, with larger values on pressure ridges. The sound-speed profile at the array site was measured several times during the experiment and proved to be very stable: A representative profile is shown in Fig. 3. A number of measurements were carried out to determine the ocean-bottom geoacoustic properties at the array site. These included a shallow sediment core,³¹ a seismic refraction survey using the hydrophone array and explosive sources in the water column,³² and a sea-floor reflectivity study which used local ice cracking as sources.³³ These measurements indicated that the seabed consisted of a layer of sand and silty clay approximately 15-m thick with a compressional-wave speed of 1800 m/s and a density of 1.8 g/cm³. Below this layer was a second bottom layer with a compressional-wave speed of 2000 m/s and a density of 1.9 g/cm³. Shear-wave speeds were taken to be 300 and 500 m/s for the first and second layers, respectively, and compressional- and shear-wave attenuations were taken to be 0.5 and 0.25 dB/λ (these values were based on a compilation of bottom properties given by Chapman).³⁴

The data analyzed here consist of a set of six 2-min samples of the ambient noise recorded over a period of approximately 2 1/2 days. The times of recording for the six samples are given in Table III. The temperature varied from –40° to –14 °C during this time. Five of the samples, denoted samples B–F, were dominated by distant ice-ridging noise. One sample, sample A, was dominated by local ther-

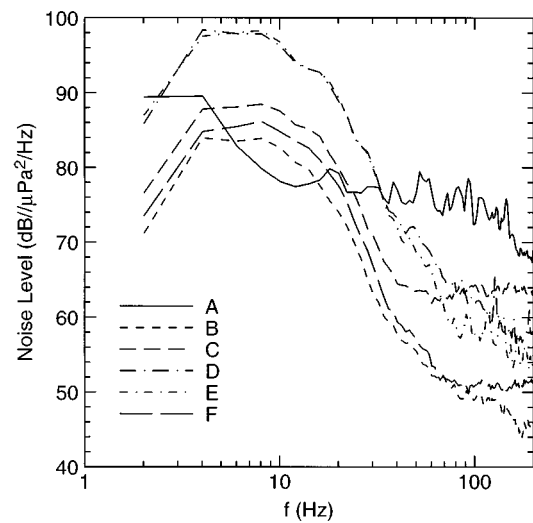


FIG. 14. Power spectra of the measured ambient noise samples. Sample A was dominated by local thermal ice cracking, samples B–F were dominated by distant ice-ridging noise.

mal ice cracking (this sample was included in the analysis for comparative purposes). The noise type was determined using spectral content, visual and audio inspection of the time series, and correlation with environmental parameters such as air temperature and wind speed.²⁸ The power spectra of the six ambient noise samples are shown in Fig. 14. Sample A exhibits the flat spectral shape typical of thermal ice cracking.^{25–28} Samples B–F exhibit a broad peak centered at approximately 10 Hz, which is typical of ice-ridge generated noise.^{17,29,30}

B. Inversion results

The parameters used for the simulated annealing inversion of the measured ambient noise fields were the same as those described for the synthetic test case in Sec. II A. No assumptions were made regarding the spectral shape of the sources (i.e., the spectral shape of each source was included in the inversion). The matching fields were calculated using the $N \times 2D$ adiabatic normal mode model described in Sec. I A (Perkins *et al.*³⁵ demonstrated that azimuthally dependent propagation effects are required to model the spatial properties of surface-generated noise). The environment was represented by the sound-speed profile and geoacoustic properties measured at the array site, as described in Sec. III A. Although these bottom properties may not apply over the entire search region, Hamson and Heitmeyer¹ and Dosso *et al.*³⁶ have shown that the mismatch in shallow water is much less sensitive to geoacoustic properties than to bathymetry (the latter reference specifically considered an Arctic environment). Accordingly, only the bathymetry has been incorporated in the inversion with a search interval of ± 50 m from values on the NRL bathymetry chart.¹⁴

An important factor influencing acoustic propagation in the Arctic is the scattering of acoustic energy at the rough underside of the ice. A number of models have been developed to treat under-ice scattering (e.g., Refs. 37–43). We have used the Burke–Twersky model,^{37,38} which represents ice ridges by a random distribution of elliptical half-

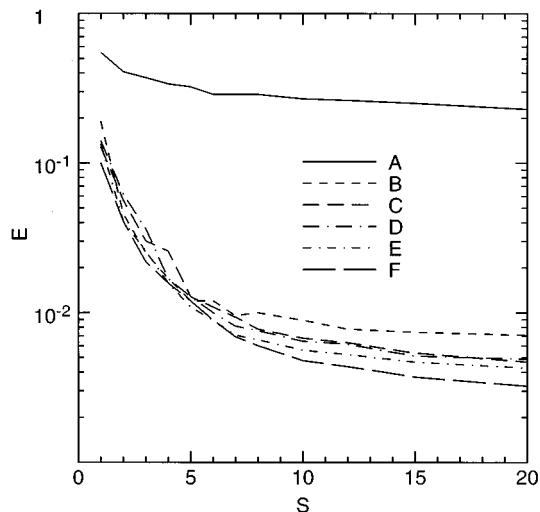


FIG. 15. Bartlett error E as a function of number of sources S sought in the inversion for noise samples A–F.

cylinders on the underside of the ice (this model is included as an option in the KRAKENC normal mode program¹¹ that was used in modeling the propagation). The Burke–Twersky model has been shown to provide good agreement with measured data in several studies;^{44–46} however, other studies have shown significant differences (e.g., Ref. 47). Since its applicability to the noise measurements considered here cannot be independently verified, under-ice scattering may be a significant source of environmental mismatch in the inversions. A ridge density of 11.5 ridges/km, with keel depths of 5.3 m and keel widths of 11.9 m, was assumed for the Burke–Twersky scattering model, based on measurements for this region reported by Diachok.³⁸ The use of other values for these parameters and/or other scattering models in the inversion could be investigated. The relative importance of bathymetry and under-ice scattering in shallow-water Arctic environments could also be studied. However, since the purpose of this example is to illustrate localization in uncertain bathymetry, we have not pursued those issues here.

Figure 15 shows the Bartlett error as a function of the number of sources sought in the inversion for the six ambient noise samples. For the samples dominated by ice-ridging noise (B–F), the mismatch decreases rapidly as the number of sources sought increases to about six or eight, then decreases much more slowly. The mismatch for seven sources is $E < 0.01$ for each of these samples, which indicates that the modeled source distributions accurately reproduce the measured acoustic data. This is illustrated in Fig. 16, which shows the magnitude of the cross-spectral matrices for noise sample E and for the modeled source distribution with seven sources (Bartlett error $E = 0.007$). All the major features of the measured cross spectra are reproduced successfully by the modeled source distribution. Thus, 6–8 sources appear to be sufficient to model the cross spectra of the ice-ridging noise samples used in this study.

For sample A, which is dominated by thermal ice cracking, the mismatch in Fig. 15 is much larger than for the ice-ridging samples and does not decrease significantly as the number of sources is increased up to 20. This indicates

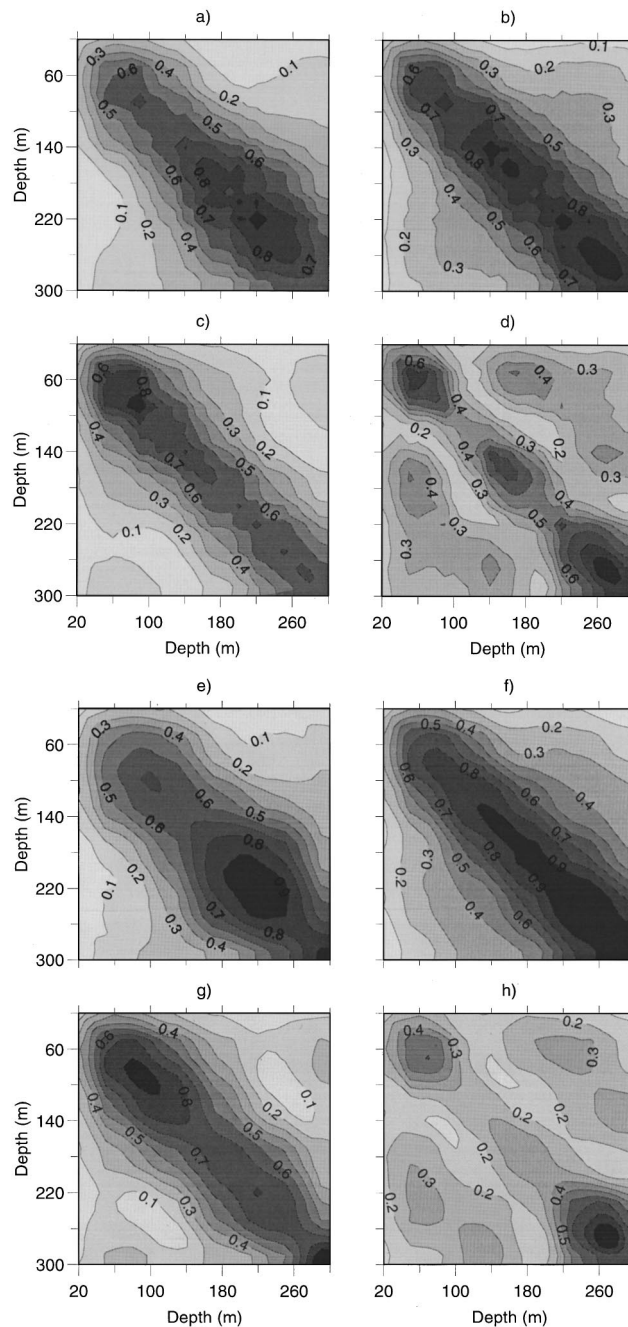


FIG. 16. Cross spectra for measured noise sample E in (a)–(d), and for the source distribution determined by inversion for seven sources in (e)–(h). Frequencies are: 10 Hz (a) and (e), 15 Hz (b) and (f), 20 Hz (c) and (g), and 30 Hz (d) and (h). The matrices are normalized so the maximum value is unity in each case. Contour lines are at intervals of 0.1 with lightest grey indicating ≤ 0.1 , and black indicating > 0.9 .

that, although the inversion models the ice-ridging noise well, the method does not have so many degrees of freedom that it can reproduce any acoustic measurements, such as local (nonmodal) ice-cracking noise.

Each of the simulated annealing inversions represented in Fig. 15 involved computing approximately 7×10^5 matching fields with the adiabatic normal mode model, and required ~ 2 h CPU time on a DEC Alpha computer (computation speed: 20 Mflops). To verify that the annealing schedule provided a sufficient search of the parameter space,

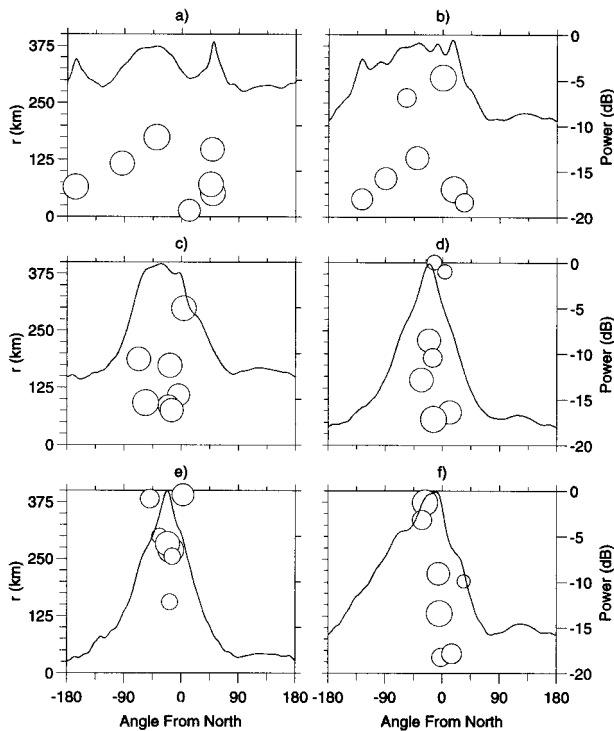


FIG. 17. Source distributions determined by inversion for samples A–F in (a)–(f), respectively. Circle size indicates relative received levels. The solid line indicates the maximum-likelihood beamforming results. True north is 0° .

several cases were repeated with a more cautious schedule that required $\sim 7 \times 10^6$ propagation models and more than 20 h of CPU time with no appreciable change in the final mismatch. To verify that the ± 50 -m search interval for the bathymetry was sufficient, several inversions were repeated with a search interval of ± 250 m, with no change in the final mismatch.

The source distributions determined by the inversions for seven sources are shown in Fig. 17. This figure also includes the results of broadband maximum-likelihood plane-wave beamforming applied to the acoustic data recorded on the horizontal array for the frequency band 10–30 Hz. The results suggest that sources located in the deep water to the north of the recording site dominate the ice-ridging samples B–F. The strength of the directivity peaks for these samples appear to evolve with time, with samples D and E exhibiting the strongest directivity. The general consistency in azimuth estimates for samples B–F suggests that the ridging sources remain aligned along constant azimuth for extended periods of time, at least several hours. The ice-cracking sample, sample A, has much flatter directivity than the ice-ridging samples, indicating an essentially isotropic distribution of sources.

Given that inverse problems are inherently nonunique, the good agreement between measured and modeled fields in Figs. 15 and 16 does not necessarily indicate that the ice-ridging sources are correctly localized. A necessary condition to claim to have correctly localized the sources would be to obtain consistent source positions for different initializations of the simulated annealing inversion algorithm. Unfortunately, this has not proven to be the case for the measured

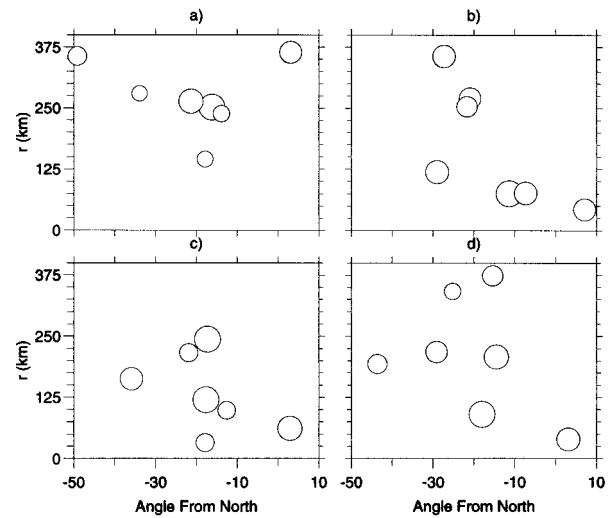


FIG. 18. Localization estimates for four initializations of the inversion algorithm for noise sample E. Circle size indicates relative received levels.

ambient noise data. Figure 18 shows the results of four inversions with different initializations applied to sample E, assuming seven sources. Each solution had an error of $E \approx 0.007$. Figure 18 indicates that consistent estimates of the source ranges were not obtained. There are a number of possible reasons for this ambiguity in range. The source distributions in Figs. 17 and 18 suggest a number of sources closely spaced in bearing. Also, the source spectral shape was not assumed known, but was included in the inversions. The synthetic study in Sec. II D indicated that in both of these cases it was more difficult to determine source ranges unambiguously. Thus, improved results could be expected in cases where the noise sources are not concentrated at small azimuthal separations or where the source spectral shape is known. Other factors which may contribute to the ambiguity in range include inadequacies of the Burke–Twersky ice-scattering model, mode coupling and three-dimensional propagation effects which are not treated by the $N \times 2D$ adiabatic mode model, errors in the location of the hydrophones, nonmodal noise from sources other than pressure-ridging, and uncertainty in environmental parameters other than bathymetry. Environmental parameters such as the geoacoustic or underice properties could be included in the inversion in a manner similar to the bathymetry. However, the inverse problem then becomes correspondingly more complex, and although this approach is worthy of further study, it has not been implemented in this paper. The equivalent bathymetry model determined in the process of minimizing the Bartlett error may also have the effect of correcting some of the environmental mismatch due to uncertainty in other environmental properties. The extent to which this occurs could be investigated in a synthetic study, but is not considered further here.

Figure 18 shows that the source bearing estimates, unlike source ranges, are reasonably consistent with respect to different initialization of the inversion algorithm. There is always a cluster of sources in the -15° to -20° direction, one source in 2° – 7° direction, and another source or cluster of sources between -30° and -35° . In addition, reinitializ-

ing the inversion algorithm for different numbers of sources did not change the conclusion that 6–8 sources were required to model the ice-ridging noise. Therefore, although the inversion method does not provide a unique model of the distribution of ice-ridging events (particularly source ranges), it does provide models which are consistent in number of sources and in azimuthal distribution that accurately reproduce the measured ambient noise fields. Thus, application of the matched-field inversion method to this problem represents an advance over plane-wave beamforming, which provided only an indication of the directivity of the ambient noise fields.

In general, matched-field inversion of measured data is a difficult problem, since in practical cases some uncertainty is associated with all environmental and array parameters, and numerical models do not exactly reproduce the measured propagation. Inverting measured fields for the spatial distribution of an unknown number of noise sources in a three-dimensional, uncertain environment is a particularly challenging case. However, we believe that applying inversion algorithms to measured as well as synthetic data is an important component in developing and evaluating new methods. The inversions presented in this section are shown to successfully treat a number of aspects of localization of multiple noise sources. The aspects of the inversion which are not completely successful serve to illustrate some of the limitations of the method for difficult problems. Significant work remains to be done on challenging acoustic inverse problems such as this.

IV. SUMMARY AND DISCUSSION

This paper describes a method for localizing an unknown number of broadband sources in a range- and azimuth-dependent environment in cases where properties of the environment are not well known. The method uses a modified Bartlett processor that matches the measured acoustic fields to fields computed for multiple sources. This allows simultaneous localization of multiple sources. The inversion searches for the positions (range and azimuth) and relative strengths of a specified number of sources, together with corrections to the bathymetry model along the propagation path to each source. The parameter search is carried out using the method of simulated annealing. The number of sources can be estimated by examining the Bartlett error returned by the inversion algorithm as a function of the number of sources sought in the inversion. The example used here to illustrate the method involved localizing an unknown number of active ice-ridge building events in a shallow-water Arctic environment where the bathymetry is poorly known.

A number of aspects of the inverse problem were investigated in a synthetic study which simulated surface noise sources in the Arctic. In this study, errors in bathymetry were found to degrade or preclude source localization. Including bathymetry in the inversion lead to successful localization of multiple sources when the true bathymetry lay within the assumed parameter search space. If the search space did not include the true bathymetry, sources may not be correctly localized. This difficulty can be overcome by increasing the

search interval for the bathymetry; however, this broadens the minima in the Bartlett mismatch function (i.e., degrades resolution). Thus, accurate bathymetry estimates are desirable.

In synthetic cases involving multiple sources, matching the acoustic data to fields produced by a single source was found to often result in incorrect source localization. Simultaneous localization was found to provide accurate positions and relative strengths for multiple sources, provided the sources were separated in azimuth and the shape of the source spectrum was known. Cases involving a number of sources at small azimuthal separation or where the spectral shape was included in the inversion yielded generally good results; however, source ranges were sometimes ambiguous and weaker sources were not always detected.

The inversion method was also applied to a set of Arctic ambient noise measurements collected in the Lincoln Sea to investigate the spatial distribution of active ice-ridge building events. Five ambient noise samples dominated by distant ice-ridging noise were analyzed; a sixth sample dominated by local thermal ice cracking was also considered for comparative purposes. Because water depths in the Arctic are often poorly known, the bathymetry in the inversions was allowed to vary by ± 50 m from published values. Source spectral shape was also included in the inversion. For the five ice-ridging noise samples, the Bartlett error was found to decrease rapidly as the number of sources in the inversion was increased to 6–8, then decreased much more slowly. All major features of the measured cross-spectral matrices were reproduced by the modeled source distribution of seven sources. Thus, 6–8 sources appear to provide a good model for the measured ice-ridging noise. For the ice-cracking noise sample, large Bartlett errors were obtained for up to 20 sources. This indicates that, although the ice-ridging noise was modeled well, the inversion method does not have so many degrees of freedom that it can reproduce any data set.

The source distributions obtained for the five ice-ridging noise samples indicate a general consistency in azimuth, suggesting that the sources remain aligned along constant azimuths for at least several hours. The source azimuths and the number of sources required were found to remain essentially constant when the inversions were repeated with different initializations. However, the source ranges were ambiguous, i.e., distributions with different source ranges were found which produced essentially the same Bartlett mismatch. This is consistent with the synthetic study which indicated that in cases where the source spectral shape is not known or where multiple sources occur at small azimuthal separation (both are the case here), source ranges tend to be ambiguous. Other factors which may contribute to the ambiguity in range include improper modeling of under-ice scattering, mode-coupling and three-dimensional propagation effects, uncertainty in environmental parameters other than bathymetry, noise from sources other than pressure-ridging, and errors in the location of the hydrophones. Future work could address the uncertainty in environmental parameters such as geoaoustic or under-ice properties by including these parameters in the inversion in the same manner as bathymetry.

- ¹R. M. Hamson and R. M. Heitmeyer, "Environmental and system effects on source localization in shallow water by the matched-field processing of a vertical array," *J. Acoust. Soc. Am.* **86**, 1950–1959 (1989).
- ²D. R. Del Balzo, D. Feuillade, and M. M. Rowe, "Effects of water-depth mismatch on matched field localization in shallow water," *J. Acoust. Soc. Am.* **83**, 2180–2185 (1988).
- ³M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," *J. Acoust. Soc. Am.* **90**, 1410–1422 (1991).
- ⁴S. E. Dosso, "Matched-field inversion for source localization with uncertain bathymetry," *J. Acoust. Soc. Am.* **94**, 1160–1163 (1993).
- ⁵J. M. Ozard, "Matched field processing in shallow water for range, depth, and bearing determination: Results of experiment and simulation," *J. Acoust. Soc. Am.* **86**, 744–753 (1989).
- ⁶A. N. Mirkin and L. H. Sibul, "Maximum likelihood estimation of the locations of multiple sources in an acoustic waveguide," *J. Acoust. Soc. Am.* **95**, 877–888 (1994).
- ⁷M. D. Collins, L. T. Fialkowski, W. A. Kuperman, and J. S. Perkins, "The multi-valued Bartlett processor and source tracking," *J. Acoust. Soc. Am.* **97**, 235–241 (1995).
- ⁸E. K. Westwood, "Broadband matched-field source localization," *J. Acoust. Soc. Am.* **91**, 2777–2789 (1992).
- ⁹M. V. Greening, "Analysis of Arctic acoustic data; modeling and prediction of Arctic ambient noise characteristics II," DREP contractor Report No. 94–110, 1994 (unpublished).
- ¹⁰A. H. Nuttall and J. H. Wilson, "Estimation of the acoustic field directionality by use of planar and volumetric arrays via the Fourier series method and the Fourier integral method," *J. Acoust. Soc. Am.* **90**, 2004–2019 (1991).
- ¹¹M. B. Porter, "The KRAKEN normal mode program," SACLANT Undersea Research Center Technical Report No. SM-245, 1991 (unpublished).
- ¹²S. Kirkpatrick, C. D. Gelatt, and M. Vecchi, "Optimization by Simulated Annealing," *Science* **220**, 671–680 (1983).
- ¹³W. A. Kuperman, M. D. Collins, J. S. Perkins, and N. R. Davis, "Optimal time-domain beamforming with simulated annealing including application of *a priori* information," *J. Acoust. Soc. Am.* **88**, 1802–1810 (1990).
- ¹⁴Naval Research Laboratory, *Bathymetry of the Arctic Ocean*, compiled by R. K. Perry and H. S. Fleming, Map and chart series MC-56 (Geological Society of America, Boulder, CO, 1986).
- ¹⁵C. A. Zala and J. M. Ozard, "Comparison of parabolic equation and adiabatic mode propagation models for matched field processing in range-dependent environments," *Proceedings of the 6th IEEE Workshop on Statistical Signal and Array Processing*, Victoria, Canada (IEEE, New York, 1992), pp. 259–262.
- ¹⁶M. D. Collins, "Benchmark calculations for higher-order parabolic equations," *J. Acoust. Soc. Am.* **87**, 1535–1538 (1990).
- ¹⁷B. M. Buck and J. H. Wilson, "Nearfield noise measurements from an Arctic pressure ridge," *J. Acoust. Soc. Am.* **80**, 256–264 (1986).
- ¹⁸P. Zakarauskas, S. E. Dosso, and J. A. Fawcett, "Matched-field inversion for source location and optimal equivalent bathymetry," *J. Acoust. Soc. Am.* **100**, 1493–1500 (1996).
- ¹⁹J. Tsao and B. D. Steinberg, "Reduction of sidelobe and speckle artifacts in microwave imaging: the CLEAN technique," *IEEE Trans. Antennas Propag.* **36**, 543–556 (1988).
- ²⁰C. A. Zala, J. M. Ozard, and M. J. Wilmut, "Prewhitening for improved detection by matched-field processing in ice-ridging correlated noise," *J. Acoust. Soc. Am.* **98**, 2726–2734 (1992).
- ²¹A. R. Milne and J. H. Ganton, "Ambient noise under Arctic sea ice," *J. Acoust. Soc. Am.* **36**, 855–863 (1964).
- ²²A. R. Milne, "Thermal tension cracking in sea ice: A source of underice noise," *J. Geophys. Res.* **77**, 2177–2192 (1972).
- ²³P. J. Stein, "Acoustic monopole in a floating ice plate," Ph.D. thesis, Woods Hole Oceanographic Institution, Woods Hole, MA, 1986.
- ²⁴D. M. Farmer and Y. Xie, "The sound generated by propagating cracks in sea ice," *J. Acoust. Soc. Am.* **85**, 1489–1500 (1989).
- ²⁵P. Zakarauskas and J. M. Thorleifson, "Directionality of ice cracking events," *J. Acoust. Soc. Am.* **89**, 722–734 (1991).
- ²⁶P. Zakarauskas, C. J. Parfitt, and J. M. Thorleifson, "Automatic extraction of spring-time Arctic ambient noise transients," *J. Acoust. Soc. Am.* **90**, 470–474 (1991).
- ²⁷M. V. Greening, P. Zakarauskas, and R. I. Verrall, "Vertical directivity of ice cracking," *J. Acoust. Soc. Am.* **92**, 1022–1030 (1992).
- ²⁸M. V. Greening and P. Zakarauskas, "Spatial and source level distributions of ice cracking in the Arctic Ocean," *J. Acoust. Soc. Am.* **95**, 783–790 (1994).
- ²⁹R. S. Pritchard, "Arctic Ocean background noise caused by ridging of sea ice," *J. Acoust. Soc. Am.* **75**, 419–427 (1984).
- ³⁰M. V. Greening and P. Zakarauskas, "Pressure ridging spectrum level and a proposed origin of the infrasonic peak in Arctic ambient noise spectra," *J. Acoust. Soc. Am.* **95**, 791–797 (1994).
- ³¹L. A. Mayers and J. Marsters, "Measurements of geophysical properties of Arctic sediment cores," DREP Contractors Report No. 89–19, Defence Research Establishment Pacific, Victoria, B.C., Canada, 1989 (unpublished).
- ³²J. P. Todoeschuch, J. M. Ozard, and J. M. Thorleifson, "Refraction and reflection experiments with a vertical array of hydrophones in the Lincoln Sea," *Eos* **69**, 1320 (1988).
- ³³P. Zakarauskas, R. I. Verrall, and M. V. Greening, "Extraction of the seabed reflectivity function using ice cracking noise as signal source," *J. Acoust. Soc. Am.* **94**, 3352–3357 (1993).
- ³⁴D. M. F. Chapman, "Surface-generated noise in shallow water: a model," *Proc. Inst. Acoust.* **9**, 1–11 (1987).
- ³⁵J. S. Perkins, W. A. Kuperman, F. Ingenito, and L. T. Fialkowski, "Modeling ambient noise in three-dimensional ocean environments," *J. Acoust. Soc. Am.* **93**, 739–752 (1993).
- ³⁶S. E. Dosso, M. L. Jeremy, J. M. Ozard, and N. R. Chapman, "Estimation of ocean-bottom properties by matched-field inversion of acoustic field data," *IEEE J. Ocean. Eng.* **18**, 232–239 (1993).
- ³⁷J. E. Burke and V. Twersky, "Scattering and reflection by elliptically striated surfaces," *J. Acoust. Soc. Am.* **40**, 883–895 (1966).
- ³⁸O. I. Diachok, "Effects of sea-ice ridges on sound propagation in the Arctic Ocean," *J. Acoust. Soc. Am.* **59**, 1110–1120 (1976).
- ³⁹D. F. Gordon and H. P. Buckner, "Model for arctic propagation loss using scattering integrals and empirical reflection losses," *J. Acoust. Soc. Am. Suppl.* **1** **75**, S74 (1984).
- ⁴⁰W. A. Kuperman and H. Schmidt, "Rough surface elastic wave scattering in a horizontally stratified ocean," *J. Acoust. Soc. Am.* **79**, 1767–1777 (1986).
- ⁴¹F. DiNapoli and R. Mellen, "Low frequency attenuation in the Arctic ocean," in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York, 1986), pp. 387–395.
- ⁴²J. R. Frike, "Acoustic scattering from elemental ice features: Numerical modeling results," *J. Acoust. Soc. Am.* **93**, 1329–1334 (1993).
- ⁴³K. LePage and H. Schmidt, "Modeling of low-frequency transmission loss in the central Arctic," *J. Acoust. Soc. Am.* **96**, 1783–1795 (1994).
- ⁴⁴J. W. Wolf, O. I. Diachok, T. C. Yang, and S. Wales, "Very-low-frequency under-ice reflectivity," *J. Acoust. Soc. Am.* **93**, 1329–1334 (1993).
- ⁴⁵T. J. Hayward and T. C. Yang, "Low-frequency Arctic reverberation. I: Measurement of under-ice backscattering strengths from short-range direct path returns," *J. Acoust. Soc. Am.* **93**, 2517–2523 (1993).
- ⁴⁶T. C. Yang and T. J. Hayward, "Low-frequency Arctic reverberation. II: Measurement of ice and bottom backscattering strengths from medium-range bottom-bounce returns," *J. Acoust. Soc. Am.* **93**, 2524–2534 (1993).
- ⁴⁷E. Livingston and O. Diachok, "Estimation of average under-ice reflection amplitudes and phases using matched-field processing," *J. Acoust. Soc. Am.* **85**, 1909–1919 (1989).

Source localization in noisy and uncertain ocean environments

Laurie T. Fialkowski, Michael D. Collins, and John S. Perkins

Naval Research Laboratory, Washington, DC 20375

W. A. Kuperman

Scripps Institution of Oceanography, La Jolla, California 92093

(Received 15 April 1996; accepted for publication 20 November 1996)

Interference from noise and uncertainties in the environmental parameters are arguably the two most serious limitations in matched-field processing (MFP). Among the techniques that have been developed for handling these difficulties are the noise-canceling processor [M. D. Collins, N. C. Makris, and L. T. Fialkowski, "Noise cancellation and source localization," *J. Acoust. Soc. Am.* **96**, 1773–1776 (1994)] and focalization [M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," *J. Acoust. Soc. Am.* **90**, 1410–1422 (1991)]. The noise-canceling processor is a generalization of the Bartlett processor that is based on matching the covariance matrix of the data with replica covariance matrices of the signal and the noise. Simulations are presented to illustrate the performance of the noise-canceling processor when there are errors in the noise replica. Focalization is a generalization of MFP in which environmental parameters are included along with source parameters in the search space. An implementation of this approach that is suitable for applications is developed and tested. The noise-canceling processor and focalization are used to simulate the localization of a source buried in noise in an uncertain environment. © 1997 Acoustical Society of America. [S0001-4966(97)03206-2]

PACS numbers: 43.60.Gk [JLK]

INTRODUCTION

Matched-field processing (MFP) is the inverse problem of localizing an acoustic source in the ocean by comparing data with solutions of the wave equation.^{1,2} MFP is very effective under ideal conditions because each propagation path contains information about the source location. Since departure from ideal conditions tends to be the rule rather than the exception, MFP has been an active area of research for nearly two decades.^{3,4} Since ocean environments are complex, there are several difficulties that arise in MFP that are not encountered in other signal-processing problems (such as beamforming). MFP techniques have been designed to handle difficulties such as environmental mismatch^{5–8} (uncertainties in the environmental parameters), interference from ambient noise and jammers,^{9,10} source motion,^{11–17} and multiple sources.¹⁷

In this paper, we implement and test MFP techniques for localizing a source buried in ambient noise, a source in an uncertain environment, and a source buried in noise in an uncertain environment. In Sec. I, we describe an implementation of focalization⁸ that is suitable for applications. In Sec. II, we describe an implementation of the noise-canceling processor¹⁰ that is suitable for applications and derive a special noise-canceling processor that is useful for many applications. In Sec. III, we investigate the performance of the noise-canceling processor as a function of signal-to-noise ratio (SNR), environmental mismatch, and noise mismatch (uncertainties in the noise covariance), and apply the noise-canceling processor and focalization to localize a source buried in noise in an uncertain environment.

I. FOCALIZATION

Environmental uncertainty (or mismatch) results in erroneous replica fields and is a serious difficulty in MFP.^{18–21} This problem is common because ocean environments are large and dynamic. Relatively small uncertainties can render MFP techniques useless even when other factors (such as the SNR) are favorable. An obvious approach for overcoming mismatch is to measure the environmental parameters accurately. This approach is usually impractical because the regions of interest are relatively large. Even small regions can be difficult to characterize when bottom interaction is significant because sediment parameters are difficult to obtain. Another approach for overcoming mismatch is to develop MFP techniques that are relatively tolerant of uncertainties.⁵

When uncertainties are large, it is necessary to include both environmental and source parameters (and possibly other parameters, such as corrections to the locations of the receivers) in the search space and perform focalization.⁸ Focalization involves adjusting the environmental parameters until the source location comes into focus in the ambiguity surface. Since the purpose of the original investigation of focalization was simply to test feasibility, the original implementations were designed to be efficient and capture the essential characteristics of the problem with little concern for applicability to data. Focalization proved to be not only feasible but also robust at determining source position due to a parameter hierarchy in which source parameters outrank environmental parameters. Focalization parameter searches often recover the source position without converging to the correct environmental parameters. This can be an advantage if the main goal is to localize the source. The promising results of the feasibility test motivated us to design an implementation that is suitable for applications.

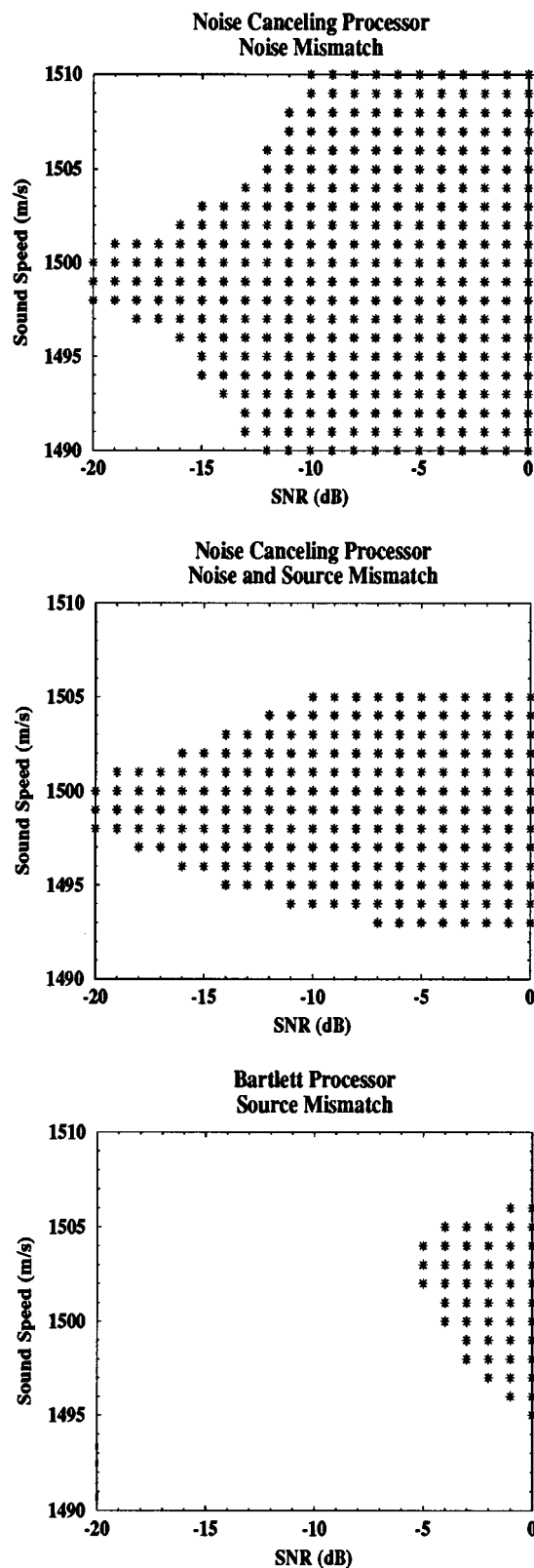


FIG. 1. Results for example A, which tests the sensitivity of the noise-canceling processor to SNR and mismatch in the noise and source replicas. The asterisks indicate combinations of mismatch and SNR for which the noise-canceling and Bartlett processors provide an estimate of the source location that lies within 600 m in range and 6 m in depth of the true source location. The mismatch parameter is the sound speed at the ocean surface.

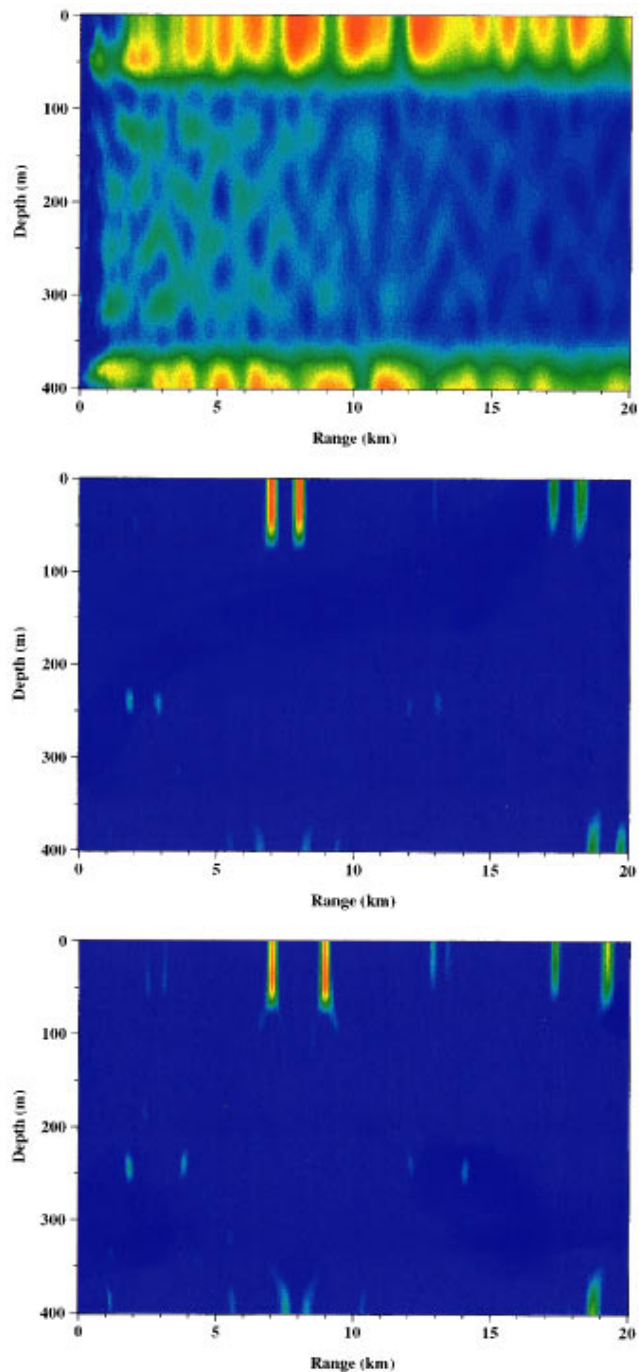


FIG. 2. Results for example B, which involves an SNR of -20 dB, a moving source, and data-based noise canceling. The source is at $r=7$ km when the noise replica is obtained. The source is at $r=8$ km and $r=9$ km when the data samples are obtained. Red corresponds to the most likely source locations. Blue corresponds to the least likely source locations. The Bartlett processor (top) contains peaks near the ocean surface because the surface-generated noise dominates the signal from the source at $r=8$ km. The noise-canceling ambiguity surfaces corresponding to $r=8$ km (middle) and $r=9$ km (bottom) contain two distinct peaks corresponding to the source positions.

Focalization involves a parameter space, cost function, and optimization algorithm. The most common MFP problems that have been considered involve two unknown parameters, the range and depth of the source. Focalization prob-

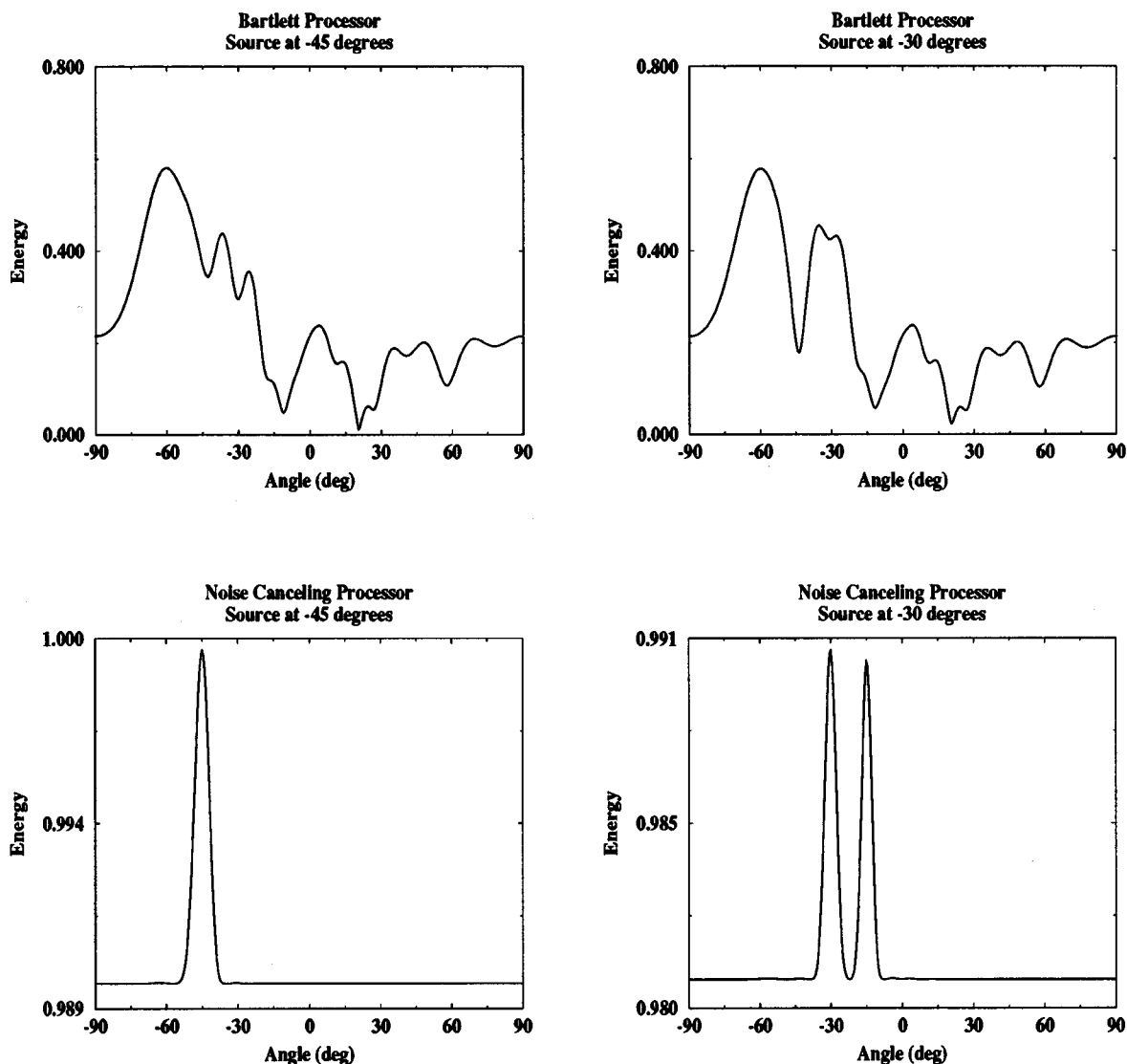


FIG. 3. Results for example C, which involves plane-wave beamforming with a towed array for a source buried by 10 dB in noise generated by the tow ship. The Bartlett processor, which appears in the top row for two cases, is dominated by the tow ship signal. The noise-canceling processor appears in the bottom row. When the source is at $\theta = -45$ deg (left column), the noise-canceling processor clearly resolves the source bearing. When the source moves from $\theta = -15$ deg to $\theta = -30$ deg (right column), data-based noise canceling clearly resolves both source bearings.

lems involve these parameters plus the focusing parameters that describe the environment and other aspects of the problem. The dimension of the parameter space may be minimized by exploiting the parameter hierarchy and including only key environmental parameters. The coordinate rotation approach of Ref. 22 can be used to identify the key parameters. The sharp peaks in parameter landscapes associated with high-resolution processors make them ineffective cost functions for focalization. We use low-resolution processors because of their simplicity and robustness. Optimization for MFP problems is often based on exhaustively searching the cost function. Since this approach is not practical for focalization problems that involve high-dimensional parameter spaces, we apply simulated annealing^{23–25} to search the focusing subspace. To avoid difficulties associated with the extremely multi-modal nature of the cost function in range-depth slices of the parameter space, we exhaustively search the source coordinates for every sample of the focusing pa-

rameters. This approach is practical because the source subspace is of low dimension.

II. THE NOISE-CANCELING PROCESSOR

In this section we describe the noise-canceling processor, a practical approach for applying it under conditions that frequently arise in practice, and a special noise-canceling processor for a special case. The data are assumed to be dominated by n temporally uncorrelated processes so that the $m \times m$ covariance matrix K is of the form

$$K = \sum_{j=1}^n K_j, \quad (1)$$

where K_j is the covariance matrix corresponding to the j th process and m is the number of receivers. We assume that the processes may be regarded as stationary over appropriate time intervals. We normalize the covariance matrices, place

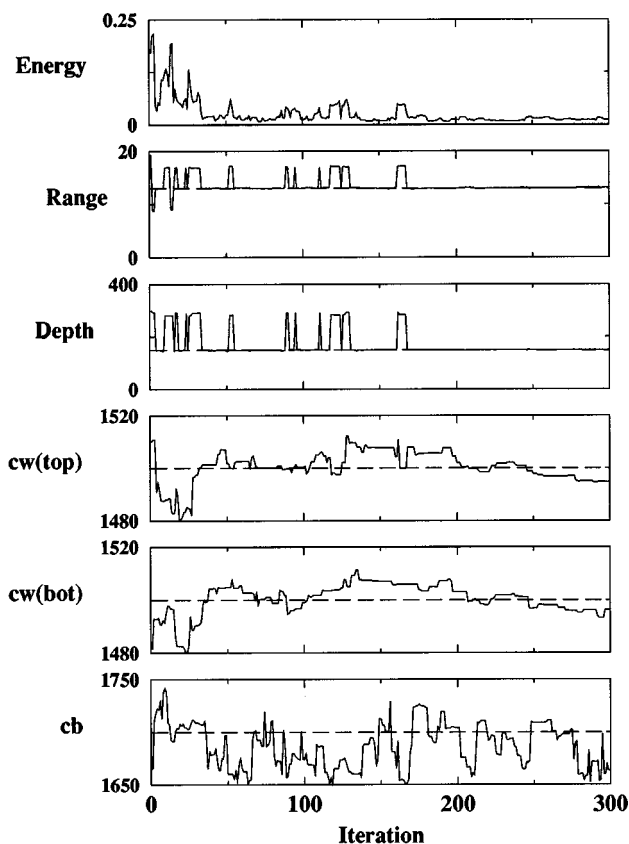


FIG. 4. The focalization parameter search for example D, which involves a source buried by 10 dB in ambient noise and three sound speed parameters. The range and depth of the source are recovered. Although the environmental parameters are not recovered, there is a strong correlation between the first two sound speed parameters. The dashed lines indicate the true parameter values.

their entries into unit vectors of dimension m^2 , and define the cost function,

$$E = \left| \mathbf{u} - \sum_{j=1}^n e_j \mathbf{u}_j \right|^2, \quad (2)$$

where the unit vector \mathbf{u} corresponds to K , the unit vector \mathbf{u}_j corresponds to K_j , and the energy levels e_j are unknown. The \mathbf{u}_j are either modeled or estimated from data.

We define A to be the $m^2 \times n$ matrix whose j th column is \mathbf{u}_j so that Eq. (2) becomes

$$E = |\mathbf{u} - A\mathbf{e}|^2, \quad (3)$$

where \mathbf{e} contains the entries e_j . The minimum of E occurs for

$$\mathbf{e} = (A^*A)^{-1}A^*\mathbf{u}. \quad (4)$$

Substituting this solution into Eq. (3), we obtain

$$B^2 = 1 - \min_e E = \mathbf{u}^* Q \mathbf{u}, \quad (5)$$

where $Q = A(A^*A)^{-1}A^*$. For the case $n=1$, the noise-canceling processor B reduces to the Bartlett processor. An estimate for the source location and other parameters is obtained by maximizing B . For this MFP problem, K corresponds to the data and the K_j correspond to the replicas.

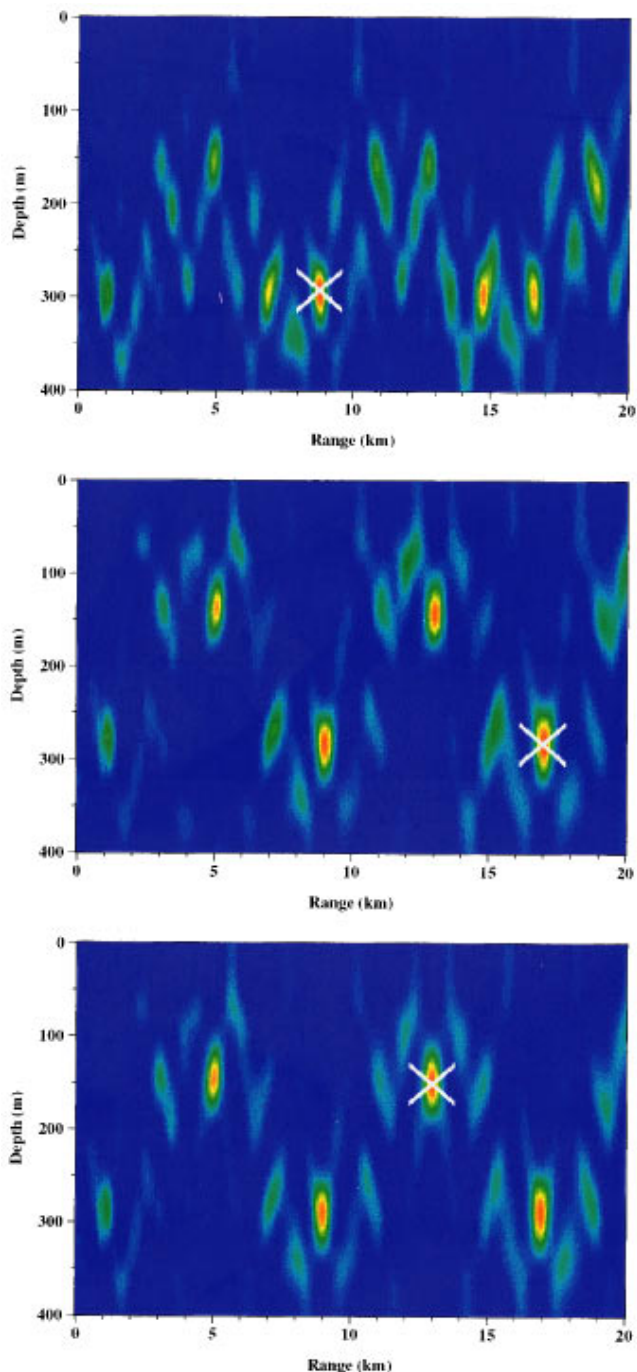


FIG. 5. Noise-canceling ambiguity surfaces encountered during the focalization parameter search for example D (from top to bottom: iterations 3, 26, and 195). Red corresponds to the most likely source locations. Blue corresponds to the least likely source locations. The crosses mark the locations of the peaks.

Since it is difficult to model noise covariance matrices in complex environments, it may be necessary to use data for the replica noise covariance matrix in practice. We refer to this approach as data-based noise canceling. In some cases, the measured noise covariance matrix may be corrupted by the signal. This difficulty can be overcome by exploiting source motion. For the case $n=2$, Eq. (2) is of the form,

$$E = |\mathbf{u}_N + \mathbf{u}_S - e_1(\mathbf{u}'_N + \mathbf{u}'_S) - e_2 \mathbf{u}_2|^2, \quad (6)$$

where $\mathbf{u} = \mathbf{u}_N + \mathbf{u}_S$ and $\mathbf{u}_1 = \mathbf{u}'_N + \mathbf{u}'_S$ are data samples obtained when the source is at different locations, \mathbf{u}_N and \mathbf{u}'_N are the contributions of the noise, \mathbf{u}_S and \mathbf{u}'_S are the contributions of the signal, and \mathbf{u}_2 is the source replica.

Assuming that the minimum of E occurs when $e_1 \cong 1$ and $\mathbf{u}'_N \cong \mathbf{u}_N$ so that the noise terms in (6) approximately cancel, we obtain

$$\min E \cong \min |\mathbf{u}_S - \mathbf{u}'_S - e_2 \mathbf{u}_2|^2. \quad (7)$$

If not for the minus sign in front of \mathbf{u}'_S , the right side of Eq. (7) would be equivalent to the Bartlett processor applied to the data $\mathbf{u} = \mathbf{u}_S + \mathbf{u}'_S$ corresponding to two sources. The minus sign does not make a significant difference if the source moves far enough so that \mathbf{u}_S and \mathbf{u}'_S are uncorrelated. In this case, the ambiguity surface is similar to the Bartlett ambiguity surface for two sources. The ambiguous peak due to the corrupting signal can be rejected by requiring that $e_2 > 0$. Stationarity is a key assumption in data-based noise canceling, which breaks down when the noise covariance varies rapidly (e.g., when the noise is dominated by nearby surface ships that are moving).

The noise-canceling processor is applicable when any or all of the K_j depend on the unknown parameters. For many problems of interest, one of the K_j depends on the unknown parameters (the variable replica) and the other K_j are independent of the unknown parameters (the fixed replicas). We derive a special noise-canceling processor for this case by interchanging the roles of the data and the variable replica so that the variable replica is matched with a combination of the data and the fixed replicas (for the general case, the data is matched with a combination of the variable and fixed replicas). We let \mathbf{pp}^* correspond to the variable replica, where \mathbf{p} contains the replica complex pressures on the array. Interchanging the roles of the data and the fixed replica in Eq. (5) and using the definitions of A and \mathbf{u} , we obtain the special noise-canceling processor,

$$B_s = |Q_s \mathbf{b}|, \quad (8)$$

where $Q_s = (A^* A)^{-1/2}$ and the j th entry of \mathbf{b} is $\mathbf{p}^* K_j \mathbf{p}$.

For the case $n = 1$, B_s reduces to the Bartlett processor. Evaluating B_s involves the evaluation of n Bartlett processors and requires slightly more than n times the computational effort that is required to evaluate the Bartlett processor. The matrix Q_s is independent of the variable replica and must be computed only once. In contrast, the matrix Q in Eq. (5) depends on the variable replica and must be computed for every sample point in the ambiguity surface. Another advantage of B_s is that the ambiguity function is normalized with respect to the replica so that $B_s \cong 1$ always indicates a good match. Since B is normalized with respect to the data, $B \cong 1$ does not necessarily indicate a good match when the SNR is low.

III. TEST PROBLEMS

In this section we apply the noise-canceling processor and focalization to test problems involving simulated data. We work in cylindrical coordinates, where the range r is the horizontal distance from an array of receivers, θ is the bearing, and z is the depth below the ocean surface. For simplic-

ity, we consider range-independent examples and use idealized models of noise mismatch. We model surface-generated noise using the approach of Ref. 26. We solve the examples using the general noise-canceling processor given by Eq. (5).

The purpose of example A is to test the performance of the noise-canceling processor when there are errors in the estimates of the noise covariance matrix and the environmental parameters. The 25-Hz source is located at $r = 13$ km and $z = 150$ m. We take $m = 13$ for the vertical array, with the j th receiver placed at $z = (-15 + 30j)$ m. The sound speed is 1500 m/s in the 400-m-deep water column. In the sediment, the sound speed is 1700 m/s, the density is 1.5 g/cm³, and the attenuation is 0.5 dB/ λ . To simulate environmental mismatch, we treat the sound speed in the water column as a linear function of depth, with the value at the surface unknown and the value at the bottom assumed to be 1500 m/s. All of the other environmental parameters are treated as known. We also use this mismatch parametrization to simulate noise mismatch. Although noise mismatch would be due to factors such as finite sampling and fluctuations in practice, the actual form of the mismatch model is probably not critical for the sake of testing.

We constructed the noise-canceling and Bartlett processors for various values of the SNR and the mismatch parameter and searched for the location of the main peak in each of the ambiguity surfaces. Mismatch errors were simulated by using the mismatch parameters to generate replicas. We considered cases in which the error is confined to the noise replica and cases in which there are errors in both the noise and source replicas. The performance of the processors is illustrated in Fig. 1, which indicates the SNR and mismatch values for which the main peak is within 600 m in range and 6 m in depth of the true source location. The Bartlett processor breaks down when the mismatch exceeds about 5 m/s or the SNR falls below about -5 dB. When there are errors in the noise replica, the noise-canceling processor is successful for much lower SNR, even when the mismatch is significant. The noise-canceling processor also performs well when there are errors in both the noise and the source replicas. The results of example A suggest that the errors in measured noise covariance matrices will be tolerable in many cases.

The purpose of example B is to illustrate data-based noise canceling. This example involves the environment, array, and frequency of example A, a moving source at $z = 25$ m, covariance matrices corresponding to three source positions, and an SNR of -20 dB. The first covariance matrix, which is used as a replica for the noise, was obtained when the source was at $r = 7$ km. The other covariance matrices, which are treated as the data, were obtained when the source was at $r = 8$ km and $r = 9$ km. Noise-canceling and Bartlett ambiguity surfaces appear in Fig. 2 for example B. The source peak is obscured by the noise in the Bartlett ambiguity surface. Since the noise replica is contaminated by a source signal, there are two prominent peaks in the noise-canceling ambiguity surfaces. Although the estimate for the source location is ambiguous in the individual noise-canceling ambiguity surfaces, the initial and final positions of the source are easy to deduce from the combination of the ambiguity surfaces.

The purpose of example C is to illustrate the application of the noise-canceling processor to plane-wave beamforming problems. This example involves a 50-Hz source and the environment of example A, with the exception that the water column is 500 m deep. We take $m=15$ for the towed horizontal line array at $z=15$ m, with the j th receiver placed at $r=(75+15j)$ m. Tow ship noise corresponding to an SNR of -10 dB is modeled using a point source at $r=0$ and $z=5$ m. Mismatch in the tow ship noise is modeled by assuming that the sediment sound speed is 1700 m/s for the data and 1850 m/s for the replica. We apply the noise-canceling processor to estimate the bearing of a distant source, which produces a signal that is approximated by a plane wave across the array. Results for example C appear in Fig. 3 for a case involving a source at $\theta=-45$ deg and a case involving a source that moves from $\theta=-15$ deg to $\theta=-30$ deg (the bearing change may be regarded as true motion or the effect of varying the heading of the tow ship). The plane-wave signal is buried in the noise in the Bartlett processor for both cases. The bearing of the plane-wave signal is clearly resolved for the fixed source. Data-based noise canceling clearly resolves the bearings of both plane-wave sources for the other case.

The purpose of example D is to illustrate the performance of focalization with the noise-canceling processor. This example involves the environment, array, source frequency, and source location of example A. We assume an SNR of -10 dB and include mismatch in the noise covariance corresponding to a sound speed of 1505 m/s at the ocean surface. As the results in Fig. 1 indicate, the noise-canceling processor is successful and the Bartlett processor fails for this combination of SNR and noise mismatch when the environmental parameters are known exactly. We parametrize the environment in terms of a constant sound speed in the sediment and a linearly varying sound speed in the water column defined by the values at $z=0$ and $z=400$ m. The density and attenuation in the sediment are assumed to be known. The water column sound speeds are assumed to lie between 1480 m/s and 1520 m/s. The sediment sound speed is assumed to lie between 1650 m/s and 1750 m/s. The noise replica is not varied during the search.

The focalization parameter search for example D is illustrated in Fig. 4. The energy decreases in the typical irregular fashion of simulated annealing. The estimates for the source range and depth converge to the correct values after sampling other locations that correspond to relatively low energy. Although the estimates for the environmental parameters do not converge, the water column parameters follow correlated paths. This behavior suggests that the efficiency of the search could be improved using the coordinate rotation approach of Ref. 22. Some of the noise-canceling ambiguity surfaces appear in Fig. 5. The parameter search attempts to bring the source into focus at two false locations before settling on the correct location.

IV. CONCLUSION

Techniques for localizing a source in noisy and uncertain environments have been tested using simulations. The noise-canceling processor outperforms the Bartlett processor

when there is mismatch in the noise and the environment. The noise-canceling processor was tested for plane-wave beamforming for a source buried in tow ship noise. Data-based noise canceling is effective when the noise replica is contaminated by a source. An implementation of focalization based on the covariance matrix is suitable for applications. Focalization was implemented using the noise-canceling processor to localize a source buried in noise in an uncertain environment. There is a need to perform further tests of focalization and noise canceling using data.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

- ¹H. P. Buckner, "Use of calculated sound fields and matched-field detection to locate sound sources in shallow water," *J. Acoust. Soc. Am.* **59**, 368–373 (1976).
- ²A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," *J. Acoust. Soc. Am.* **83**, 571–587 (1988).
- ³A. 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).
- ⁴M. D. Collins and W. A. Kuperman, "Inverse problems in ocean acoustics," *Inverse Probl.* **10**, 1023–1040 (1994).
- ⁵H. Schmidt, A. B. Baggeroer, W. A. Kuperman, and E. K. Scheer, "Environmentally tolerant beamforming for high-resolution matched field processing: Deterministic mismatch," *J. Acoust. Soc. Am.* **88**, 1851–1862 (1990).
- ⁶A. M. Richardson and L. W. Nolte, "A *a posteriori* probability source localization in an uncertain sound speed, deep ocean environment," *J. Acoust. Soc. Am.* **89**, 2280–2284 (1991).
- ⁷J. L. Krolik and W. S. Hodgkiss, "Matched field source localization in an uncertain environment using constraints based upon sound speed perturbations," in *Proceedings of the IEEE Oceans '91 Conference* (IEEE, New York, 1991).
- ⁸M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," *J. Acoust. Soc. Am.* **90**, 1410–1422 (1991).
- ⁹H. B. Riley and J. A. Tague, "Matched-field localization in high noise environments: A reduced-rank signal processing approach," *J. Acoust. Soc. Am.* **96**, 1515–1520 (1994).
- ¹⁰M. D. Collins, N. C. Makris, and L. T. Fialkowski, "Noise cancellation and source localization," *J. Acoust. Soc. Am.* **96**, 1773–1776 (1994).
- ¹¹H. C. Song and A. B. Baggeroer, "The resolution of modal Doppler shifts in a dispersive oceanic waveguide," *J. Acoust. Soc. Am.* **88**, 268–282 (1990).
- ¹²E. K. Westwood, "Broadband matched-field source localization," *J. Acoust. Soc. Am.* **91**, 2777–2789 (1992).
- ¹³C. A. Zala and J. M. Ozard, "Matched-field processing for a moving source," *J. Acoust. Soc. Am.* **92**, 403–417 (1992).
- ¹⁴J. A. Fawcett and B. H. Maranda, "A hybrid target motion analysis/matched-field processing localization method," *J. Acoust. Soc. Am.* **94**, 1363–1371 (1993).
- ¹⁵B. H. Maranda and J. A. Fawcett, "Localization of a maneuvering target using simulated annealing," *J. Acoust. Soc. Am.* **94**, 1376–1384 (1993).
- ¹⁶M. D. Collins, L. T. Fialkowski, W. A. Kuperman, and J. S. Perkins, "Environmental source tracking," *J. Acoust. Soc. Am.* **94**, 3335–3341 (1993).
- ¹⁷M. D. Collins, L. T. Fialkowski, W. A. Kuperman, and J. S. Perkins, "The multivalued Bartlett processor and source tracking," *J. Acoust. Soc. Am.* **97**, 235–241 (1995).
- ¹⁸D. R. Del Balzo, C. Feuillade, and M. M. Rowe, "Effects of water-depth mismatch on matched-field localization in shallow water," *J. Acoust. Soc. Am.* **83**, 2180–2185 (1988).
- ¹⁹A. Tolstoy, "Sensitivity of matched field processing to sound-speed profile mismatch for vertical arrays in a deep water Pacific environment," *J. Acoust. Soc. Am.* **85**, 2394–2404 (1989).
- ²⁰D. F. Gingras, "Methods for predicting the sensitivity of matched-field processors to mismatch," *J. Acoust. Soc. Am.* **86**, 1940–1949 (1989).

- ²¹R. M. Hamson and R. M. Heitmeyer, "Environmental and system effects on source localization in shallow water by the matched-field processing of a vertical array," J. Acoust. Soc. Am. **86**, 1950–1959 (1989).
- ²²M. D. Collins and L. Fishman, "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**, 1637–1644 (1995).
- ²³N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1091 (1953).
- ²⁴S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).
- ²⁵H. Szu and R. Hartley, "Fast simulated annealing," Phys. Lett. **122**, 157–162 (1987).
- ²⁶W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).

Quantitative measures of hair cell loss in CBA and C57BL/6 mice throughout their life spans

Vlasta P. Spongr

Hearing Research Laboratory, Department of Communicative Disorders and Sciences, 215 Parker Hall,
SUNY at Buffalo, Buffalo, New York 14214

Dorothy G. Flood^{a)}

Department of Neurology, University of Rochester, School of Medicine and Dentistry, Rochester,
New York 14642

Robert D. Frisina

Otolaryngology Division, Surgery Department, University of Rochester, School of Medicine and Dentistry,
Rochester, New York 14642-8629

Richard J. Salvi

Hearing Research Lab, 215 Parker Hall, SUNY at Buffalo, Buffalo, New York 14214

(Received 5 November 1996; revised 24 January 1997; accepted 28 January 1997)

The CBA mouse shows little evidence of hearing loss until late in life, whereas the C57BL/6 strain develops a severe and progressive, high-frequency sensorineural hearing loss beginning around 3–6 months of age. These functional differences have been linked to genetic differences in the amount of hair cell loss as a function of age; however, a precise, quantitative description of the sensory cell loss is unavailable. The present study provides mean values of inner hair cell (IHC) and outer hair cell (OHC) loss for CBA and C57BL/6 mice at 1, 3, 8, 18, and 26 months of age. CBA mice showed little evidence of hair cell loss until 18 months of age. At 26 months of age, OHC losses in the apex and base of the cochlea were approximately 65% and 50%, respectively, and IHC losses were approximately 25% and 35%. By contrast, C57BL/6 mice showed approximately a 75% OHC and a 55% IHC loss in the base of the cochlea at 3 months of age. OHC and IHC losses increased rapidly with age along a base-to-apex gradient. By 26 months of age, more than 80% of the OHCs were missing throughout the entire cochlea; however, IHC losses ranged from 100% near the base of the cochlea to approximately 20% in the apex. © 1997 Acoustical Society of America.

[S0001-4966(97)04606-7]

PACS numbers: 43.64.Dw, 43.64.Wn [RAS]

INTRODUCTION

Because of the varying hearing abilities of different strains and its relatively short life span, the mouse has become a popular animal model for physiological and anatomical studies of age-related hearing loss (Ehret, 1975; Ehret *et al.*, 1983; Willott, 1984, 1986; Li and Borg, 1991; Willott *et al.*, 1993; Trune *et al.*, 1996). For example, there are interesting genetic differences which cause some strains, such as the C57BL/6, to lose their hearing early in life whereas other strains, such as the CBA, maintain relatively normal hearing into senescence (Willott, 1986; Willott *et al.*, 1988a, b; Shone *et al.*, 1991; Li and Borg, 1992; Walton *et al.*, 1995). In addition, some strains of mice which show a loss of auditory sensitivity early in life also appear to be more susceptible to noise-induced hearing loss (Shone *et al.*, 1991; Erway *et al.*, 1996).

Genetic differences among different strains of mice can influence the time course and pattern of cochlear pathology. Some strains, such as the CBA, show few signs of degeneration in the organ of Corti and appear to retain much of their

hearing function well into the second year of life (Willott, 1986; Willott *et al.*, 1988b). However, in very old CBA mice (30 months), there is a mild loss (~25%–30%) of spiral ganglion cells which is relatively uniform along the length of the cochlea (Willott, 1991; Willott and Mortenson, 1991). Several strains of mice show a preferential loss of OHCs early in life followed by a subsequent loss of IHCs (Henry and Chole, 1980).

One strain of mouse which has been extensively studied physiologically and anatomically is the C57BL/6 strain. Previous studies have shown that sensory hair cells in the C57BL/6 strain begin to degenerate relatively early in life (Mikaelian, 1979; Henry and Chole, 1980; Shone *et al.*, 1991; Willott, 1991; Kazee *et al.*, 1995). In addition, there is a pronounced base-to-apex gradient of spiral ganglion cell loss during the first 12 months of life which progresses to a severe loss of ganglion cells by 30 months of age (Willott, 1991).

Because the auditory periphery of the CBA mouse appears to remain relatively intact until later life, it has become a popular model for studying the effects of aging on the central auditory system, i.e., a relatively “intact ear connected to an old brain” (Willott, 1986; Willott *et al.*, 1991). On the other hand, the C57BL/6 mouse, which shows con-

^{a)}Current address: Cephalon, Inc., 145 Brandywine Parkway, West Chester, PA 19380.

siderable loss of cochlear hair cells and spiral ganglion neurons early in life, has often been used to study the effects of cochlear pathology on the function of the central nervous system, i.e., “a damaged ear connected to a young brain” (Willott, 1986, 1991; Parham and Willott, 1988; Kazee *et al.*,

1995; Walton *et al.*, 1995). Despite the fact that the C57BL/6 and CBA mice have become extremely popular animal models for studying the effects of aging and peripheral hearing loss on central auditory system function, we are unaware of quantitative data showing the time course and pattern of hair

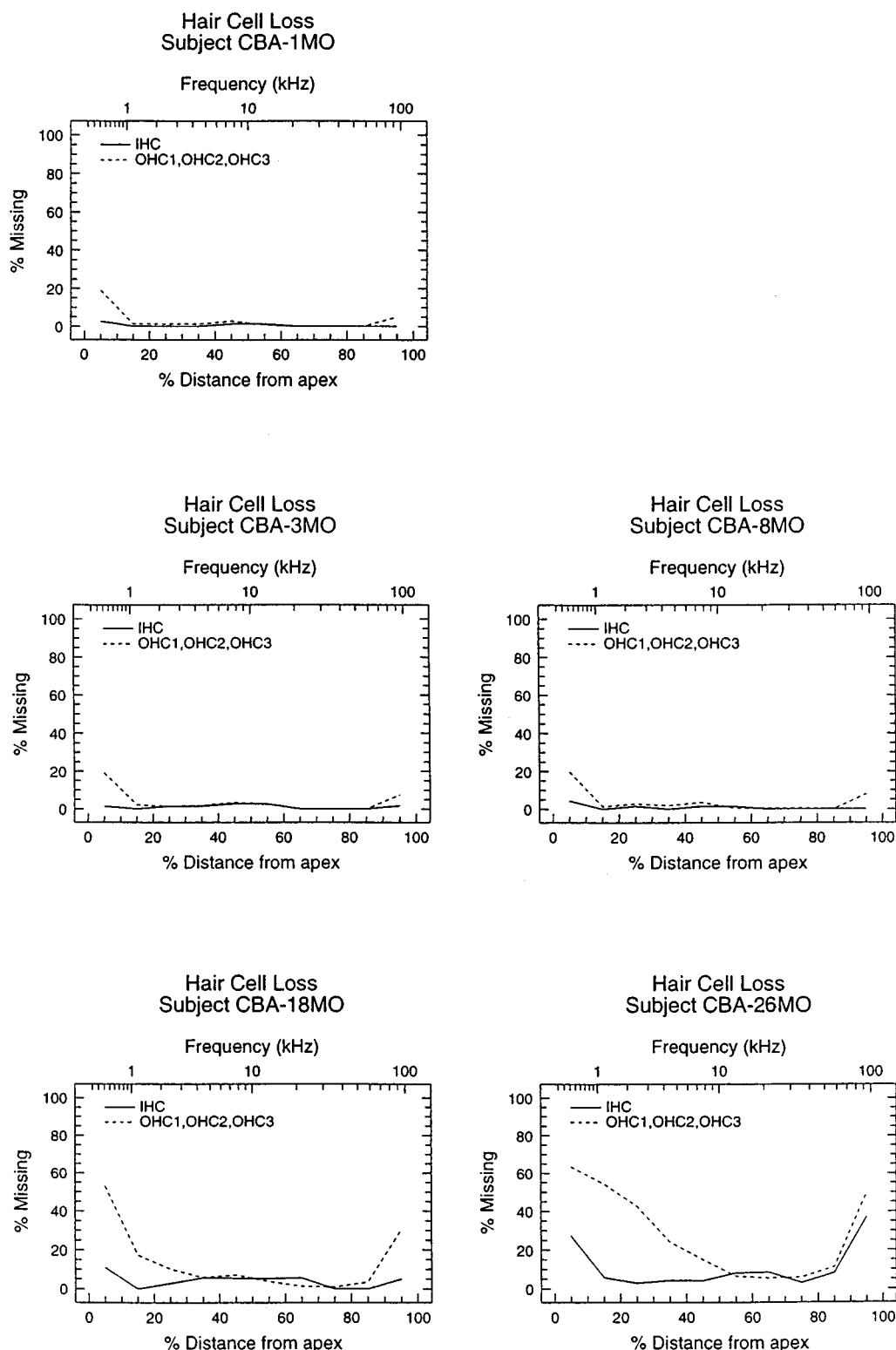


FIG. 1. Mean cytochrome loss for CBA mice at 1 ($n=10$), 3 ($n=10$), 8 ($n=5$), 18 ($n=10$), and 26 ($n=10$) months of age. Cytochrome loss shows percentage of missing OHCs and IHCs in 10% intervals as a function of percent total distance from the apex to the base of the cochlea; cochlear position was transformed to frequency using a generalized cochlear frequency-place map conversion equation for the mouse (Greenwood, 1990).

cell loss as a function of age in these two mouse strains. A quantitative description of the pattern of hair cell loss as a function of age could prove useful in interpreting the existing anatomical data on spiral ganglion cell loss (Willott, 1991) as well as the wealth of physiological data from the peripheral and central auditory system (Willott, 1986; Willott *et al.*, 1991, 1993; Kazee *et al.*, 1995; Walton *et al.*, 1995). The purpose of the present study was to provide a quantitative description of the percentage of IHC and OHC loss as a function of age in the CBA and C57BL/6 strains.

I. METHODS

A. Subjects

Five age groups of C57BL/6 mice and five age groups of CBA mice were used as subjects. Approximately half of the C57BL/6 mice were purchased from the Charles River Breeding Laboratory; the remainder were obtained from Harlan Sprague Dawley. CBA mice were from two strains; approximately two-thirds were CBA/Ca, purchased from Charles River Breeding Laboratory, and one-third were CBA/J, purchased from Harlan Sprague Dawley. Most of the mice were purchased and then kept in a room in the University of Rochester animal facility for one week to two months before being sacrificed for histological analysis. The mean ages for the five groups were: 1 month, 3 months, 8 months, 18 months, and 26 months. Ten cochleas were evaluated for each age group except for the 8 month old, CBA group. Since there was minimal hair cell loss between 1 month and

8 months of age in the CBA group, only 5 cochleas were examined for the 8 month group. All subjects had normal external and middle ears at the time of sacrifice.

B. Methods

Our methods for preparing and analyzing the cochleas have been described previously (Boettcher *et al.*, 1992; Spongr *et al.*, 1992). In this study, mice were deeply anesthetized with sodium pentobarbital (35–80 mg/kg) and perfused transcardially with 0.1% heparin sulfate and 0.5% sodium nitrite in 0.85% sodium chloride for 10 min. This was followed by perfusion with either 4% paraformaldehyde or by 1% glutaraldehyde and 4% paraformaldehyde in phosphate buffer (pH 7.3) for 25 min (Kazee *et al.*, 1995). The cochleas were subsequently removed from the temporal bone and the round window and oval window membrane were opened. Some cochleas were transferred to phosphate buffered saline and stored at 4 °C until further processing. Some cochleas were again fixed by perfusion of 2.5% glutaraldehyde in veronal acetate buffer (pH 7.2) through the round window and then immersed in the same fixative at 4 °C until further processing. The cochleas were rinsed in buffer, post-fixed in 1% osmium tetroxide for 1 h, and dehydrated up to 70% ETOH. The bone surrounding the cochleas was thinned and the cochleas subsequently perfused with 0.35 M EDTA. The decalcified cochleas were microdissected and half-turns of each organ of Corti were removed, trimmed, and mounted in glycerin on glass slides (Spongr *et al.*, 1992). The sensory

TABLE I. % hair cell loss statistics in CBA mice ($n = 10/\text{age group}$).^a

| Age in months (mean) | | 1 | | 3 | | 8 | | 18 | | 26 | |
|---------------------------|---------|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| % Distance | Row | Mean | s.d. | Mean | s.d. | Mean | s.d. | Mean | s.d. | Mean | s.d. |
| 1–10 | OHC 1-3 | 18.75 | 6.85 | 19.56 | 11.16 | 19.79 | 11.31 | 52.83 | 10.78 | 63.78 | 18.45 |
| | IHC | 2.78 | 1.88 | 0.98 | 9.19 | 4.29 | 3.13 | 11.04 | 3.53 | 27.81 | 30.79 |
| 11–20 | OHC 1-3 | 1.13 | 1.45 | 2.37 | 3.00 | 1.49 | 2.04 | 17.76 | 7.23 | 54.43 | 23.71 |
| | IHC | 0.00 | 0.04 | 0.00 | 0.60 | 0.00 | 1.09 | 0.66 | 3.85 | 5.24 | 8.14 |
| 21–30 | OHC 1-3 | 1.50 | 1.00 | 1.58 | 1.34 | 2.99 | 1.54 | 10.21 | 6.19 | 42.83 | 24.96 |
| | IHC | 0.26 | 0.51 | 0.98 | 1.25 | 0.88 | 0.78 | 2.81 | 2.48 | 3.25 | 2.47 |
| 31–40 | OHC 1-3 | 1.28 | 1.72 | 1.53 | 2.94 | 1.87 | 2.19 | 5.70 | 4.50 | 23.83 | 18.00 |
| | IHC | 0.06 | 0.42 | 0.80 | 1.41 | 0.56 | 0.77 | 5.97 | 3.60 | 4.04 | 3.83 |
| 41–50 | OHC 1-3 | 3.08 | 1.11 | 3.65 | 2.94 | 3.23 | 1.34 | 7.28 | 4.58 | 15.23 | 10.80 |
| | IHC | 1.63 | 0.03 | 3.23 | 2.76 | 2.04 | 0.63 | 6.06 | 2.98 | 4.25 | 3.34 |
| 51–60 | OHC 1-3 | 0.99 | 1.36 | 2.12 | 4.32 | 0.44 | 0.61 | 3.75 | 2.51 | 6.46 | 5.80 |
| | IHC | 0.87 | 0.04 | 3.01 | 2.10 | 1.51 | 0.82 | 5.91 | 3.39 | 7.48 | 4.23 |
| 61–70 | OHC 1-3 | 0.00 | 1.84 | 0.00 | 1.68 | 0.44 | 2.91 | 1.29 | 2.78 | 5.13 | 5.93 |
| | IHC | 0.00 | 0.06 | 0.00 | 1.70 | 0.00 | 1.49 | 5.25 | 2.77 | 8.46 | 10.24 |
| 71–80 | OHC 1-3 | 0.00 | 0.41 | 0.00 | 0.53 | 0.36 | 3.83 | 1.12 | 2.14 | 5.48 | 4.82 |
| | IHC | 0.00 | 0.09 | 0.00 | 0.66 | 0.00 | 0.81 | 0.00 | 3.68 | 3.44 | 3.93 |
| 81–90 | OHC 1-3 | 0.00 | 0.05 | 0.00 | 0.42 | 0.18 | 1.92 | 3.50 | 3.68 | 11.07 | 8.11 |
| | IHC | 0.00 | 0.07 | 0.00 | 0.19 | 0.00 | 0.84 | 0.65 | 1.72 | 7.42 | 8.31 |
| 91–100 | OHC 1-3 | 4.81 | 6.68 | 7.07 | 10.30 | 7.41 | 8.49 | 30.09 | 26.68 | 49.69 | 26.04 |
| | IHC | 0.00 | 0.04 | 2.30 | 4.34 | 0.55 | 2.76 | 5.80 | 4.04 | 36.39 | 24.12 |
| Length of cochleas (mean) | | 5.91 | 0.31 | 5.58 | 0.33 | 5.29 | 0.60 | 5.59 | 0.53 | 5.59 | 0.23 |

^a $n = 5/\text{age group}$ 8.

epithelium was viewed with a differential interference contrast microscope (Zeiss Standard) at a magnification of $1000\times$ (oil immersion). Counts of the number of IHCs and OHCs, present or missing, were obtained over 0.1–0.18-mm intervals and entered into a personal computer using custom software that allowed averaging of the data from several animals. Sensory cells were counted as present if the cell body and cuticular plate were intact and as missing if a phalangeal scar was present. Percent hair cell loss was determined on the basis of the hair cell density distribution in each mouse strain (Spongr *et al.*, 1994). Individual cochleograms were determined by plotting the percentage of missing IHCs and OHCs as a function of percent total distance from the apex of the cochlea. Percent distance from the apex of the cochlea was mapped to frequency using a generalized cochlear frequency-place map formula (Greenwood, 1990). The average IHC loss and average OHC loss values were determined for each 10% interval of the cochlea and a mean cochleogram was plotted using the data from all of the cochleas in each group (Figs. 1 and 3).

II. RESULTS

A. CBA strain

Figure 1 shows the average IHC and OHC loss in 10% intervals as a function of percent distance from the apex of the cochlea in 1-, 3-, 8-, 18-, and 26-month-old CBA mice. Table I shows the means and standard deviations (sd) for IHC loss and OHC loss (combined loss of all three rows). Few IHCs and OHCs were missing at 1 month of age; however, a few phalangeal scars in the OHC region were observed in the extreme basal and apical ends of the cochlea. In the base (90%–100% distance from apex), approximately 5% of the OHCs were missing and in the apex (0%–10% distance from apex), approximately 19% were absent. The average IHC and OHC losses at 3 months and 8 months of age were virtually identical to those at 1 month of age. The first obvious signs of OHC damage were observed in the extreme apex and base of the cochlea at 18 months of age. The OHC loss had increased to approximately 53% in the apical 10% (>0.7 kHz) of the cochlea and to approximately 30% in the basal 10% (<80 kHz) of the cochlea. IHC losses were typically less than 6% except in the extreme apical region where it increased to around 10%. The hair cell losses in the apical and basal regions of the cochlea showed a significant increase between 18 and 26 months of age resulting in a broad “U-shaped” cochleogram. OHC loss decreased from about 64% at the apex to around 20% near the middle of the cochlea (~ 10 kHz) and then increased again to around 50% in the basal 20% (>45 kHz) of the cochlea. IHC loss also followed a broad “U-shape” with approximately a 28% loss in the extreme apex and a 37% loss in the extreme base; in the middle of the cochlea the IHC loss was on the order of 5%–10%. At 26 months of age, the magnitude of OHC loss in the basal third of the cochlea tended to be greatest in OHC3; however, in the apex of the cochlea, the loss tended to be greatest in OHC1.

Figure 2 illustrates the growth of OHC and IHC loss as a function of age in 20% segments along the length of the

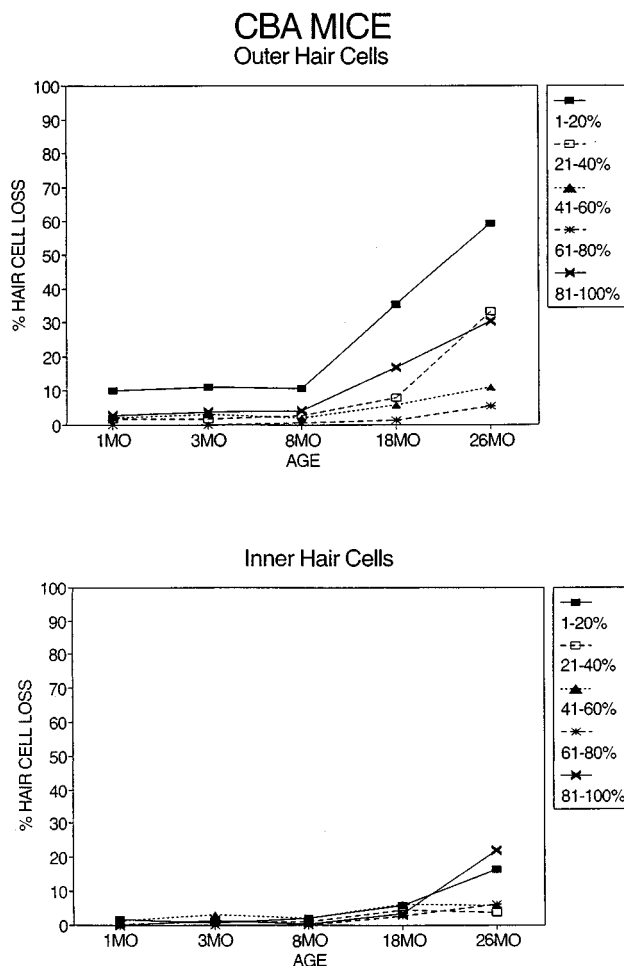


FIG. 2. Mean percentage of OHC (top) and IHC (bottom) loss in 20% segments of the cochlea of the CBA mouse as a function of age. Segments refer to percent distance from the apex of the cochlea.

cochlea. OHC losses were small ($\sim 10\%$) and showed no increase during the first 8 months of life. Between 8 and 26 months of age, OHC loss in the region 1%–21% of the distance from the apex increased from approximately 10% to 60%. During this same time period, OHC loss in more basal regions (81%–100%, and 61%–100%) increased to approximately 30%. In contrast, IHC losses were negligible during the first 18 months of life, but showed a small increase ($\sim 15\%$) in the extreme apical and basal regions between 18 and 26 months of age.

B. C57 strain

Figure 3 shows the average IHC and OHC loss as a function of percent distance from the apex of the cochlea 1-, 3-, 8-, 18-, and 26-month-old C57 mice. The means and standard deviations (sd) for the IHC loss and OHC loss are shown in Table II. Only a few ($<10\%$) OHCs were missing in the extreme basal and apical region of the 1 month old C57 mouse (Fig. 3; Table II). By 3 months of age, the OHC and IHC losses in the base of the cochlea had increased to 80% and 55%, respectively. In contrast, few hair cells ($\sim 5\%$) were missing in the apical 75% of the cochlea. By 8 months of age, OHC and IHC losses were evident through-

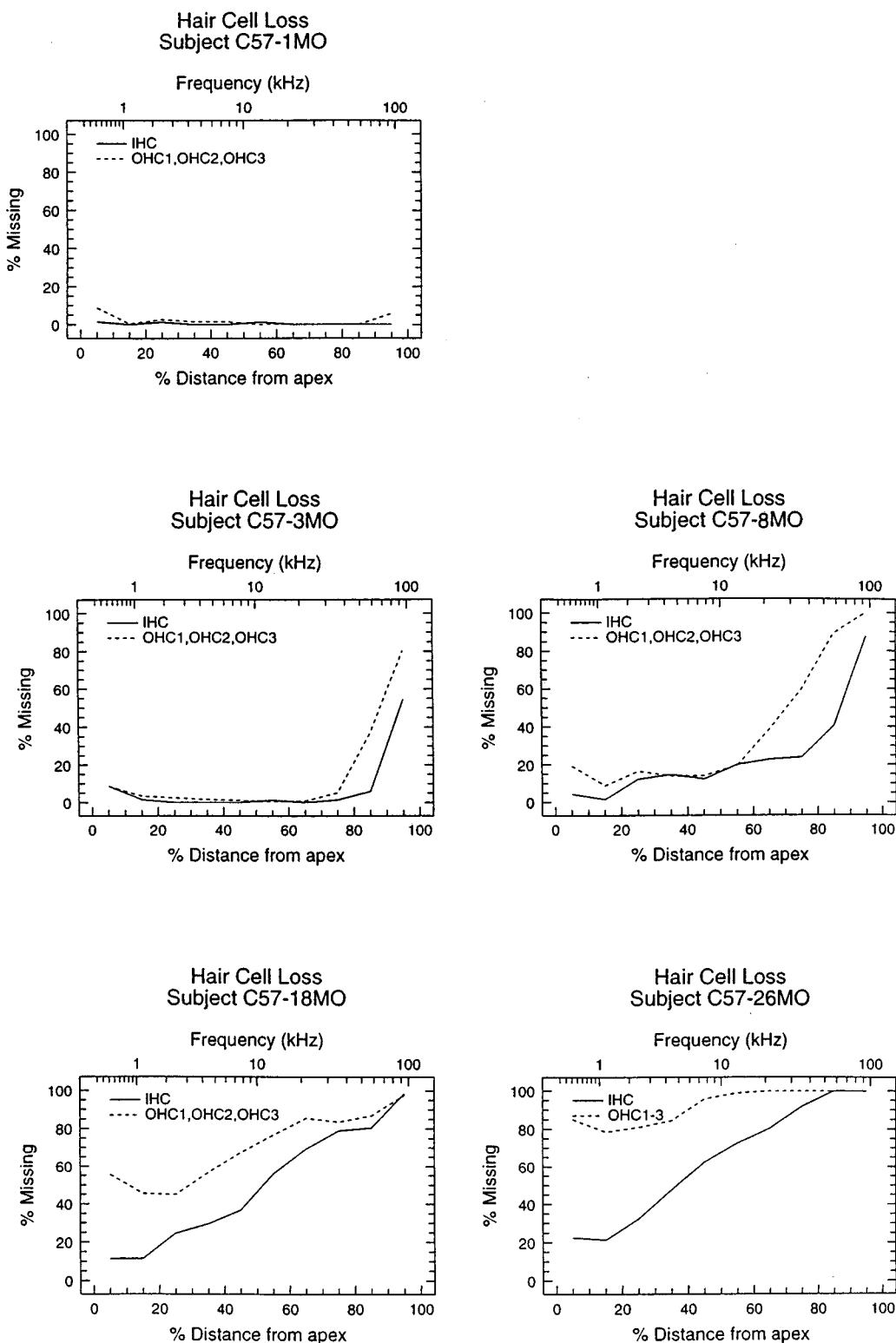


FIG. 3. Mean ($n = 10$) cytochleograms for C57BL/6 mice at 1, 3, 8, 18, and 26 months of age. Cytochleograms show percentage of missing OHCs and IHCs in 10% intervals as a function of percent total distance from the apex of the cochlea; cochlear position was transformed to frequency using a generalized cochlear frequency-place map conversion equation for the mouse (Greenwood, 1990).

out the cochlea. In the extreme basal region (~ 90 kHz), the OHC loss was on the order of 100% and then gradually decreased to around 15%–20% in the apical half of the cochlea (< 10 kHz). The IHC loss in the extreme base was on the order of 87%, but then rapidly decreased to 15%–20% in the middle two-thirds of the cochlea. The OHC and IHC

losses rapidly expanded toward the apex of the cochlea between 8 and 26 months of age. At 18 months of age, the OHC loss increased from approximately 50% in the apical third of the cochlea to more than 85% in the basal third of the cochlea. On the other hand, the IHC loss showed a steady decrease from approximately 100% in the extreme base

TABLE II. % hair cell loss statistics in C57BL/6 mice ($n=10$ /age group).

| Age in months (mean) | | 1 | | 3 | | 8 | | 18 | | 26 | |
|---------------------------|---------|------|------|-------|-------|-------|-------|-------|-------|--------|-------|
| % Distance | Row | Mean | s.d. | Mean | s.d. | Mean | s.d. | Mean | s.d. | Mean | s.d. |
| 1–10 | OHC 1-3 | 8.48 | 3.60 | 9.36 | 8.33 | 19.31 | 3.76 | 58.93 | 12.07 | 84.61 | 20.21 |
| | IHC | 2.07 | 2.12 | 8.66 | 6.82 | 4.86 | 3.48 | 11.35 | 5.95 | 22.26 | 17.79 |
| 11–20 | OHC 1-3 | 0.76 | 1.09 | 3.23 | 2.38 | 8.78 | 13.88 | 48.69 | 18.84 | 78.21 | 24.13 |
| | IHC | 0.00 | 0.00 | 0.99 | 1.41 | 1.17 | 3.86 | 12.49 | 5.35 | 21.05 | 16.59 |
| 21–30 | OHC 1-3 | 3.72 | 3.47 | 2.64 | 2.16 | 16.57 | 27.90 | 47.54 | 24.54 | 80.70 | 27.42 |
| | IHC | 1.92 | 0.13 | 0.38 | 1.27 | 12.23 | 22.67 | 26.24 | 8.70 | 32.23 | 14.64 |
| 31–40 | OHC 1-3 | 1.54 | 1.28 | 1.72 | 1.47 | 14.04 | 29.06 | 60.99 | 28.93 | 84.53 | 29.41 |
| | IHC | 0.64 | 0.03 | 0.20 | 0.75 | 14.44 | 27.95 | 31.05 | 6.59 | 47.55 | 23.19 |
| 41–50 | OHC 1-3 | 1.19 | 1.05 | 1.28 | 2.32 | 13.96 | 28.69 | 73.02 | 30.64 | 95.47 | 9.13 |
| | IHC | 0.00 | 0.00 | 0.00 | 1.93 | 12.61 | 19.09 | 38.72 | 15.98 | 62.02 | 20.34 |
| 51–60 | OHC 1-3 | 0.43 | 1.14 | 0.50 | 3.94 | 19.28 | 31.13 | 82.66 | 31.10 | 99.18 | 2.23 |
| | IHC | 1.06 | 0.40 | 1.30 | 2.69 | 19.96 | 26.44 | 59.59 | 24.54 | 71.98 | 21.55 |
| 61–70 | OHC 1-3 | 0.19 | 0.31 | 0.70 | 3.43 | 38.25 | 41.49 | 90.39 | 28.82 | 99.95 | 0.13 |
| | IHC | 0.06 | 0.04 | 0.43 | 0.66 | 22.29 | 31.08 | 73.87 | 27.91 | 80.95 | 19.90 |
| 71–80 | OHC 1-3 | 0.17 | 0.53 | 5.64 | 8.45 | 59.40 | 29.11 | 85.85 | 30.27 | 100.00 | 0.00 |
| | IHC | 0.00 | 0.00 | 0.86 | 0.92 | 23.84 | 28.55 | 83.90 | 25.64 | 92.03 | 12.96 |
| 81–90 | OHC 1-3 | 0.00 | 0.02 | 37.11 | 26.87 | 88.92 | 19.85 | 86.01 | 28.68 | 100.00 | 0.00 |
| | IHC | 0.03 | 0.09 | 6.05 | 5.12 | 40.11 | 34.66 | 84.88 | 31.89 | 99.68 | 0.95 |
| 91–100 | OHC 1-3 | 7.50 | 4.11 | 81.56 | 31.39 | 99.92 | 0.21 | 97.06 | 7.24 | 100.00 | 0.00 |
| | IHC | 0.87 | 2.19 | 59.90 | 32.43 | 87.87 | 14.88 | 99.32 | 2.05 | 100.00 | 0.00 |
| Length of cochleas (mean) | | 6.20 | 0.20 | 6.11 | 0.47 | 6.13 | 0.32 | 6.12 | 0.15 | 6.21 | 0.35 |

of the cochlea to approximately 15% in the extreme apex. By 26 months of age, nearly 100% of the OHCs were missing in the basal half of the cochlea (>10 kHz) and more than 80% were missing in the apical half of the cochlea. The IHC loss, by contrast, showed only a slight increase over that seen at 18 months of age. The magnitude of OHC loss at 26 months of age tended to be greatest in OHC3 in the apical third of the cochlea.

Figure 4 shows the increase in hair cell loss as a function of age in the C57 mouse. Between 1 and 8 months of age, OHC loss in the region 81%–100% of the distance from the apex of the cochlea (map frequencies >45 kHz) increased rapidly to approximately a 95% loss. Between 3 and 18 months of age, OHC loss in the region of the cochlea 61%–80% of the distance from the apex of the cochlea (~ 16 –45 kHz) increased to approximately 90%. OHC losses in the apical 60% of the cochlea showed significant increases after 8 months of age. By 26 months of age OHC loss increased to approximately 96%, 85%, and 84% in cochlear regions 41%–60% (~ 6 –16 kHz), 21%–40% (~ 1.6 –6 kHz), and 1%–20% (<1.6 kHz) of the distance from the apex, respectively.

IHC losses occurred later and progressed more slowly than OHC losses. IHC loss in the extreme base of the cochlea (81%–100% region) increased from 0 to approximately 92% between 1 and 18 months of age. IHC losses in the region 61%–80% of the distance from the base increased from 0 to 79% between 3 and 18 months of age. By 26 months of age, IHC loss increased to approximately 65%, 40%, and 20% in

cochlear regions 41%–60% (~ 6 –16 kHz), 21%–40% (~ 1.6 –6 kHz), and 1%–20% (<1.6 kHz) of the distance from the apex, respectively.

III. DISCUSSION

The results of the present study provide a quantitative description of the OHC and IHC losses as a function of age in CBA and C57BL/6 mice. The pattern and time course of hair cell loss is distinctly different in these two strains (Figs. 1 and 3) (Henry and Chole, 1980; Willott and Mortenson, 1991). A few phalangeal scars were present in the extreme apical and basal regions of the cochlea of the CBA mouse at 1 month of age. The first sign of sensory cell loss in the CBA strain was seen at 18 months of age in the extreme apical and basal ends of the cochlea. These results indicate that hair cell loss develops relatively slowly between 8 and 18 months of age. Hair cell losses were more pronounced in the apex than the base and mainly involved the loss of OHCs. Our group data also suggest that hair loss begins somewhat earlier than the sample data reported by Henry and Chole (1980). By 26 months of age, more than 40% of the OHCs in the apical third of the cochlea were missing. In addition, there was a mild loss of IHCs in both the basal and apical ends of the cochlea; however, the hair cell population in the middle of the cochlea was largely intact. The pattern of hair cell loss reported here is similar to that reported by Henry and Chole (1980); however, our average OHC loss in the apex was not as large as the data shown in their study. The lack of signifi-

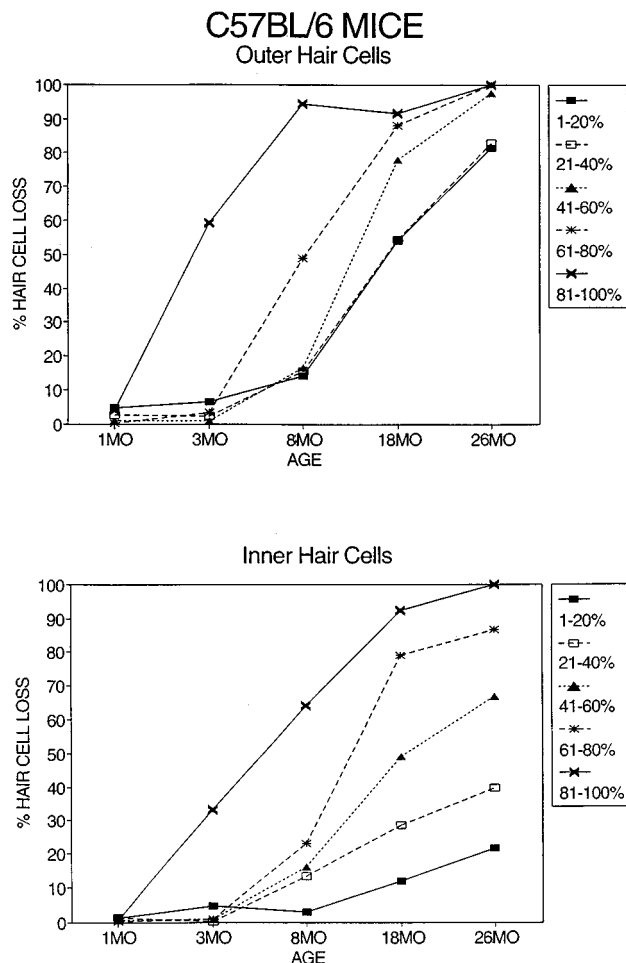


FIG. 4. Mean percentage of OHC (top) and IHC (bottom) loss in 20% segments of the cochlea of the C57BL/6 mouse as a function of age. Segments refer to percent distance from the apex of the cochlea.

cant hair cell loss in the CBA mouse until relatively late in life is similar to that reported for Sprague–Dawley rats (Keithley and Feldman, 1982), chinchilla (Bohne *et al.*, 1990) and gerbil (Schmiedt *et al.*, 1990).

Previous electrophysiological studies have shown that the CBA mouse retains relatively good hearing sensitivity through middle age. Near the end of its 2- to 2.5-year life span, the CBA mouse shows a moderate loss of sensitivity over a broad range of frequencies (Henry, 1983; Willott, 1986; Hunter and Willott, 1987). Auditory brainstem response (ABR) thresholds to high-pass and low-pass noise pips increased only about 10 dB between 1 and 19 months of age and ABR amplitudes only decreased slightly (Hunter and Willott, 1987). The small changes in the ABR at 19 months of age are consistent with the relatively small hair cell losses seen in our 18 month old CBA mice. Neural thresholds have also been assessed in the inferior colliculus of 22 month old CBA mice. The greatest loss of sensitivity, 20–30 dB, occurred between 16 and 40 kHz; neurons with best frequencies below 8 kHz showed a smaller loss (~ 10 dB) of sensitivity (Willott, 1986). The pattern of hearing loss, which has a maximum in the mid-frequencies, is inconsistent with the cochleogram which shows little or no IHC and OHC loss in the region of the cochlea (50%–80% of the distance from the

apex) corresponding to these frequencies. Given that the greatest loss of OHCs occurs in apical 40% of the cochlea, we would have expected the greatest loss of sensitivity to have occurred below 8 kHz. The reason for the discrepancy is unclear, but one possibility is that the surviving hair cells in the region of the cochlea corresponding to 16–40 kHz may be present, but functioning abnormally (Liberman and Kiang, 1978; Salvi *et al.*, 1982). In addition, ganglion cell loss is slightly greater in the middle of the cochlea than in the apex (Willott and Mortenson, 1991). A third possibility is that the cochlear frequency-place map may not be appropriate for these two strains of mice. Single neuron labeling of physiologically characterized neurons would be extremely useful in assessing the validity of the cochlear maps in CBA and C57 mice (Liberman, 1982).

In contrast to the CBA mouse, the hair cell losses in the C57 mouse rapidly expand from the base of the cochlea toward the apex (Mikaelian, 1979; Henry and Chole, 1980; Willott and Mortenson, 1991). Significant OHC and IHC losses were already evident by 3 months of age in our study. By 26 months of age, nearly all of the OHCs were missing throughout the cochlea, whereas IHC losses gradually tapered off from 100% near the base of the cochlea to around 20% loss near the apex. Previous studies have also found significant loss of spiral ganglion neurons by 12 months of age, and by 26 months of age nearly all the ganglion cells are missing from the base of the cochlea and roughly half are missing from the apex (Willott and Mortenson, 1991). The pattern of hair cell loss seen in the C57 mouse is similar to that of the A/J strain (Chole and Henry, 1983).

Previous electrophysiological studies have shown that C57 mice develop a high-frequency sensorineural hearing loss around 6–8 months of age (Mikaelian, 1979; Henry and Chole, 1980; Willott, 1986; Hunter and Willott, 1987). Previous ABR thresholds in C57 mice show a high-frequency hearing loss that extends down to 10–20 kHz (Walton *et al.*, 1995). Our cochleograms from 8 month old C57 mice show only a partial loss of OHCs and IHCs in the base of the cochlea. The region of OHC loss in our mean cochleogram corresponds to approximately 15 kHz or greater on the cochlear frequency map, whereas the region of IHC loss corresponds to frequencies of 40 kHz or greater (Greenwood, 1990). These results suggest that a significant proportion of the IHCs which are present in the basal half of the cochlea may be functioning abnormally. This interpretation is consistent with other results from acoustically traumatized ears (Liberman and Kiang, 1978; Salvi *et al.*, 1982). By 18–24 months of age, the hearing loss had spread to the lowest frequencies and the mice were almost completely deaf (Mikaelian, 1979; Henry and Chole, 1980; Willott, 1986; Hunter and Willott, 1987). Despite the fact that C57 mice have almost no residual hearing by 2 years of age, our results show that more than 50% of the IHCs are still present in the apical third of the cochlea at this time. This would suggest that the IHCs which are present at this time are extremely insensitive to sound. This large loss of sensitivity is undoubtedly partly due to the massive OHC loss.

ACKNOWLEDGMENTS

Research supported by the National Institute of Aging, NIH Grant No. P01 AG 09524 and the International Center for Hearing and Speech Research, Rochester, NY. The excellent technical assistance of Li-Ying Han and Yin Guo is greatly appreciated.

- Boettcher, F. A., Spongr, V. P., and Salvi, R. J. (1992). "Physiological and histological changes associated with the reduction in threshold shift during interrupted noise exposure," *Hearing Res.* **62**, 217–236.
- Bohne, B., Gruner, M. M., and Harding, G. W. (1990). "Morphological correlates of aging in the chinchilla cochlea," *Hearing Res.* **48**, 79–82.
- Chole, R. A., and Henry, K. R. (1983). "Disparity in the cytochrome c oxidase and the electrocochleogram in aging LP/J and A/J inbred mice," *Audiology* **22**, 384–392.
- Ehret, G. (1975). "Frequency and intensity discrimination limen and non-linearities in the ear of the house mouse (*Mus Musculus*)," *J. Comp. Physiol. Psychol.* **102**, 321–336.
- Ehret, G., Moffat, A. J. M., and Capranica, R. R. (1983). "Two-tone suppression in auditory nerve fibers of the green tree frog (*Hyla cinerea*)," *J. Acoust. Soc. Am.* **73**, 2093–2095.
- Erway, L., Shiu, Y.-W., Davis, R. R., and Krieg, E. F. (1996). "Genetics of age-related hearing loss in mice. III. Susceptibility of inbred and F1 hybrid strains to noise-induced hearing loss," *Hearing Res.* **93**, 181–187.
- Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2604.
- Henry, K. R. (1983). "Aging and audition," in *Auditory Psychobiology of the Mouse* (C. C. Thomas, Springfield), pp. 470–493.
- Henry, K. R., and Chole, R. A. (1980). "Genotypic differences in behavioral, physiological and anatomical expressions of age-related hearing loss on the laboratory mouse," *Audiology* **19**, 369–383.
- Hunter, K. P., and Willott, J. F. (1987). "Aging and the auditory brainstem response in mice with severe or minimal presbycusis," *Hearing Res.* **30**, 207–218.
- Kazee, A. M., Han, L. Y., Spongr, V. P., Walton, J. P., Salvi, R. J., and Flood, D. G. (1995). "Synaptic loss in the central nucleus of the inferior colliculus correlates with sensorineural hearing loss in the C57BL/6 mouse model of presbycusis," *Hearing Res.* **89**, 109–120.
- Keithley, E. M., and Feldman, M. L. (1982). "Hair cell counts in age-graded series of rat cochleas," *Hearing Res.* **8**, 249–262.
- Li, H. S., and Borg, E. (1991). "Age-related loss of auditory sensitivity in two genotypes," *Acta Oto-Laryngol.* **111**, 827–834.
- Li, H.-S., and Borg, E. (1992). "Influence of genotype and age on acute acoustic trauma and recovery in CBA/Ca and C57BL/6J mice," *Acta Oto-Laryngol.* **112**, 956–967.
- Lieberman, M. C. (1982). "Single-neuron labeling in the cat auditory nerve," *Science* **216**, 1239–1241.
- Lieberman, M. C., and Kiang, N. Y. S. (1978). "Acoustic trauma in cats: Cochlear pathology and auditory nerve activity," *Acta Oto-Laryngol. Suppl.* **358**, 5–63.
- Mikaelian, D. O. (1979). "Development and degeneration of hearing in the C57b16 mouse: Relation of electrophysiologic responses from the round window and cochlear nucleus to cochlear anatomy and behavioral responses," *Laryngoscope* **89**, 1–15.
- Parham, K., and Willott, J. F. (1988). "The acoustic startle response in young and aging C57BL/6J and CBA/J mice," *Behav. Neurosci.* **102**, 881–886.
- Salvi, R. J., Perry, J., Hamernik, R. P., and Henderson, D. (1982). "Relationship between cochlear pathologies and auditory nerve and behavioral responses following acoustic trauma," in *New Perspectives on Noise-Induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. J. Salvi (Raven, New York), pp. 165–188.
- Schmiedt, R. R., Mills, J. H., and Adams, J. C. (1990). "Tuning and suppression in auditory nerve fibers of aged gerbils raised in quiet or noise," *Hearing Res.* **45**, 221–236.
- Shone, G., Altschuler, R. A., Miller, J. M., and Nuttall, A. L. (1991). "The effects of noise exposure on the aging ear," *Hearing Res.* **56**, 173–178.
- Spongr, V. P., Boettcher, F. A., Saunders, S. S., and Salvi, R. J. (1992). "Effects of noise and salicylate on hair cell loss in the chinchilla cochlea," *Arch. Otolaryngol. Head Neck Surg.* **118**, 157–164.
- Spongr, V., Powers, N., Flood, D., and Salvi, R. (1994). "Difference in hair cell density distribution in the cochlea of young CBA/HSD and C57BL/HSD mice," *Abstr. Assoc. Res. Otolaryngol.* **17**, 91.
- Trune, D. R., Kempton, J. B., and Mitchell, C. (1996). "Decreased auditory function in the C3H/lpr autoimmune disease mouse," *Hearing Res.* **95**, 57–92.
- Walton, J. P., Frisina, R. D., and Meierhans, L. R. (1995). "Sensorineural hearing loss alters recovery from short-term adaptation in the C57BL/6 mouse," *Hearing Res.* **88**, 19–26.
- Willott, J. F. (1984). "Changes in frequency representation in the auditory system of mice with age-related hearing impairment," *Brain Res.* **309**, 159–162.
- Willott, J. F. (1986). "Effects of aging, hearing loss, and anatomical location on thresholds of inferior colliculus neurons in C57BL/6 and CBA mice," *J. Neurophysiol.* **57**, 391–408.
- Willott, J. F. (1991). *Aging and the Auditory System: Anatomy, Physiology, and Psychophysics* (Singular Press, San Diego).
- Willott, J. F., and Mortenson, V. (1991). "Age-related cochlear histopathology in C57BL/6J and CBA/J mice," *Abstr. Assoc., Res. Otolaryngol.* **14**, 16.
- Willott, J. F., Parham, K., and Hunter, K. R. (1988a). "Response properties of inferior colliculus neurons in middle-aged C57BL/6J mice with presbycusis," *Hearing Res.* **37**, 15–28.
- Willott, J. F., Parham, K., and Hunter, K. R. (1988b). "Response properties of inferior colliculus neurons in young and very old CBA/J mice," *Hearing Res.* **37**, 1–14.
- Willott, J. F., Parham, K., and Hunter, K. P. (1991). "Comparison of the auditory sensitivity of neurons in the cochlear nucleus and inferior colliculus of young and aging C57BL/6J and CBA/J mice," *Hearing Res.* **53**, 78–94.
- Willott, J. F., Aitkin, L. M., and McFadden, S. L. (1993). "Plasticity of auditory cortex associated with sensorineural hearing loss in adult C57BL/6J mice," *J. Comp. Neurol.* **329**, 402–411.

Fine structure of the $2f_1-f_2$ acoustic distortion product: Effects of primary level and frequency ratios

Ning-ji He^{a)} and Richard A. Schmiedt

Department of Otolaryngology and Communicative Sciences, Medical University of South Carolina,
171 Ashley Avenue, Charleston, South Carolina 29425-2242

(Received 5 August 1996; revised 13 December 1996; accepted 14 December 1996)

The fine structure of the $2f_1-f_2$ acoustic distortion product (ADP) was measured in humans with different primary level (L_1/L_2) and frequency ($f_2/f_1, f_2 > f_1$) ratios. The (L_1/L_2) ratio was varied under two conditions. In the first condition L_1 was fixed at 50 dB SPL while L_2 was varied from 30 to 75 dB SPL in 5-dB steps. An upward frequency shift was observed in the ADP fine structure as L_2 was increased. In the second condition, L_2 was fixed at 50 dB SPL and L_1 varied, and a downward frequency shift was observed. These opposing frequency shifts are predicted by a vector-sum model [Sun *et al.*, J. Acoust., Soc. Am. **96**, 2166–2174, 2175–2183 (1994)] and support the hypothesis that the ADP fine structure largely reflects place features of the area of overlap of the primary traveling waves. The mechanisms underlying the shifts in fine structure were further investigated by using three primary f_2/f_1 ratios: 1.11, 1.2, and 1.33. An orderly difference in the rate of fine-structure shift with level was observed as a function of f_2/f_1 ratio, with the largest rate of shift associated with the smallest frequency ratio. This observation, along with the fact that downward frequency shift (with L_1 varied) is always at a larger rate than the upward shift (with L_2 varied), suggests that ADP levels and fine structure are strongly influenced by the nonlinear compression present in the mechanics of the basilar membrane in the region of overlap between the primary traveling waves. © 1997 Acoustical Society of America. [S0001-4966(97)01205-8]

PACS numbers: 43.64.Jb, 43.64.Bt [RDF]

INTRODUCTION

The $2f_1-f_2$ acoustic distortion product (ADP) is now widely used in both clinical and laboratory tests of auditory function because it is frequency specific, noninvasive, and it occurs in both human and animal subjects. Unfortunately, its applications are limited somewhat by the fact that the generation mechanism for the ADP is not fully understood. Some of the most fundamental questions concerning its site of generation are still unresolved by experimental data. A current issue of debate is whether the ADP is generated within a single cochlear region, or whether it is the product of contributions from multiple regions.

An example of a multiple-region hypothesis is one where a re-emission from the $2f_1-f_2$ place interacts with emissions from the primary place. Studies of the relationship between the ADP and other kinds of emissions have shown that ADP amplitudes can be enhanced by the presence of strong spontaneous emissions (SOAEs) or stimulus frequency emissions (SFEs) associated with the frequency region near the $2f_1-f_2$ place (Wilson, 1980; Wit *et al.*, 1981; Kemp and Brown, 1984; Furst *et al.*, 1988). This enhancement has been interpreted as evidence for the presence of a re-emission from the $2f_1-f_2$ place. However, it is not clear whether the enhancement is place specific or just a result of a wave summation of the two emissions within the cochlea.

On the other hand, most experimental data have shown that a tone with a frequency close to f_2 has a dominant effect

on the ADP in noise-exposure and fatigue paradigms (Dolan and Abbas, 1985a,b; Martin *et al.*, 1987; Schmiedt, 1984, 1986), and in suppression studies (Kemp and Brown, 1984; Brown and Kemp, 1984; Martin *et al.*, 1987; Harris *et al.*, 1992). Recently, Kimberley *et al.* (1995) estimated traveling wave delays for the ADP and SFE at $2f_1-f_2$ frequencies and found that the delay for the ADP was shorter than that for SFE, suggesting a more basal source for the ADP than that for SFE. Indeed, most data suggest that the ADP originates largely within the region on the basilar membrane corresponding to the primary tones. However, the picture is not yet clear; data have been reported in the suppression literature, showing that suppression sometimes peaks at frequencies other than f_2 (Krummer *et al.*, 1995). These data seem to favor the re-emission hypothesis. Suppression studies with the ADP are complex: The third tone not only suppresses the $2f_1-f_2$ component but also interacts with f_1 or f_2 as a third stimulus tone and generates additional distortion products (Harris *et al.*, 1992).

Recently, ADP fine structures have been observed in human subjects. The fine structures are characterized by a series of sharp peaks and valleys across frequency with a peak-to-valley depth of up to 20 dB (Gaskill and Brown, 1990; He and Schmiedt, 1993). The peak-to-peak distance of the ADP fine structure is about 3/32 octave, a value consistent with the minimal frequency distance for the consecutive peaks observed in SOAEs, SFEs, and threshold fine structure (Schloth, 1983; Zwicker, 1990; Zwicker and Schloth, 1984; Zwicker and Peisl, 1990). Furthermore, a systematic downward frequency shift in the ADP fine structure with increas-

^{a)}Electronic mail: hening@musc.edu

ing levels of both (equal-level) primaries has been observed (He and Schmiedt, 1993). This shift is qualitatively and quantitatively consistent with what is found in the basilar-membrane response of chinchilla (Ruggero and Rich, 1991), suggesting that the ADP fine structure manifests some place features of the basilar membrane and serves as a promising tool to investigate basilar-membrane compressive nonlinearities and the origin of the overall ADP.

By itself, the existence of the ADP fine structure raises questions about the origins of the ADP. Some investigators believe the fine structure is evidence for multiple generators. Pioneering work by Kemp and Brown (1983) showed that the rippling pattern can be suppressed by a third tone at a frequency slightly higher than $2f_1 - f_2$. They concluded that the fine structure results from re-emissions from the $2f_1 - f_2$ place, which then summate with the basalward traveling-wave component of the ADP in a mechanism similar to that hypothesized for the SFE. However, the possibility of the suppression tone affecting more than one site cannot be ruled out, given the small frequency interval between the $2f_1 - f_2$ component and the primaries (≤ 0.58 octave for $f_2/f_1 = 1.2$) and the shallow slope of the suppression tuning curve on the low-frequency side (~ 10 dB/octave).

Similar ambiguities are often noticeable in other data examining the correlation of the ADP fine structure to other measures. Piskorski *et al.* (1995) reported that with f_2 fixed, a frequency coincidence in fine structures among different-order distortion product components (i.e., $2f_1 - f_2$, $3f_1 - 2f_2$, ...) was observed when data were plotted as a function of distortion product frequency. The data were explained in terms of the re-emission of these different components from similar characteristic places associated with the distortion products. However, because the f_2 frequency was fixed, the results can also be interpreted as a consequence of a relatively fixed generator within the primary region.

In a previous study (He *et al.*, 1994), the fine structure of the behavioral threshold was measured over a 1/3-octave frequency range and the ADP fine structure over a wider frequency range covering the frequency regions corresponding to f_2 , f_1 , and $2f_1 - f_2$. A cross-correlation analysis showed a cyclic function of the correlation, i.e., the correlation coefficient varied from highly positive to significantly negative in a period similar to that of the ADP fine structure. No consistent trend for the peaks was observed among subjects. These results suggest that caution must be used when comparing fine structures: Rippling patterns with similar peak-to-peak intervals may not necessarily be generated from the same place.

Another way of examining the mechanisms underlying fine structure is through modeling. A computer model based on a vector-sum algorithm has been developed to simulate the ADP fine structure and its level-dependent frequency shift (Sun *et al.*, 1994a,b). The model assumes that there is an array of ADP generators, taken as outer hair cells (OHCs), extending basalward from the primary regions, and that the ADP results from a vector sum of all wavelets arising from these generators. A stochastic damping factor is associated with each OHC. The damping factor is randomly distributed around a mean value. It was found that this irregularity is a

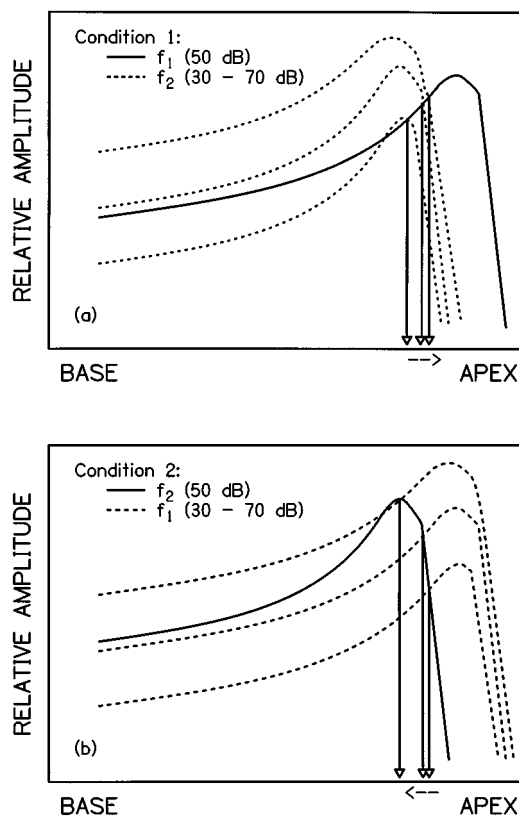


FIG. 1. Basilar-membrane excitation patterns in response to f_1 and f_2 based on a vector-sum model (Sun *et al.*, 1994a). (a) In condition 1, L_1 is fixed at 50 dB SPL and L_2 is varied from 30 to 70 dB SPL. The location of the maximum of the overlap between the primary traveling waves is indicated by the solid arrow. Note that the maximum shifts toward the apex with increasing L_2 , as indicated by the horizontal arrow at the bottom. (b) In condition 2, L_2 is fixed and L_1 is varied, and the overlap peak moves basalward with increasing L_1 . Thus, the model predicts opposing frequency shifts in the ADP fine structure under conditions 1 and 2.

necessary condition for this model to simulate the ADP fine structure. Realistic ADP fine structures are generated with this model without the necessity for any re-emissions from the $2f_1 - f_2$ place.

To study further the generation mechanism of the ADP, a series of experiments were conducted to test the model of Sun *et al.* (1994a,b). An approach was used that relies on the place features of the ADP fine structure and its basilar-membrane-related frequency shift. An advantage of this approach over more conventional suppression paradigms is that it avoids the complexities caused by the third tone (suppressor). In the experiments reported here, the effects of both the level ratio and the frequency ratio of the primaries on the ADP fine structure are investigated.

I. RATIONALE

Figure 1 presents model predictions for the envelopes of cochlear traveling waves for different L_1/L_2 ratios under two conditions. In condition 1, L_1 was fixed while L_2 was varied (upper panel, a). In condition 2, L_2 was fixed and L_1 varied (lower panel, b). According to our model, these manipulations result in different patterns of excitation along the basilar membrane in response to the primary tones. With

fixed L_1 [Fig. 1(a)] the region of peak overlap between the primary traveling waves shifts toward the apex as L_2 increases, mainly due to the expansion of the excitation pattern for f_2 . This apical shift translates to an upward frequency shift in the place characteristics for the overlap. Conversely, when L_2 is fixed [Fig. 1(b)], the region of peak overlap moves basalward, and a downward frequency shift is expected as L_1 increases. Thus, the first hypothesis to be tested is that opposing frequency shifts in the ADP fine structure are predicted for condition 1 (L_1 is fixed and L_2 varied) and condition 2 (L_2 is fixed and L_1 varied).

The model can also be tested with different frequency ratios of the primaries to investigate further the mechanism underlying the fine-structure shift with frequency. In the model the frequency shift of the traveling-wave envelope with increasing primary level is achieved by assuming a level-dependent damping constant. Higher levels increase the damping, resulting in increased compression of the response as well as a basalward shift in tuning. Increasing the effective level in the area of overlap (thereby increasing the damping) can also be accomplished by making the frequency ratio of the primaries smaller. Under these conditions, more of the overlap area is associated with the traveling-wave peaks where the nonlinear damping is most effective. Thus, the second hypothesis is that the level-dependent compression and frequency shift should be a function of the amount of overlap; i.e., increasing amounts of the overlap should result in increasing amounts of frequency shift in the ADP fine structure as a function of primary level. To test this hypothesis, three representative f_2/f_1 ratios were used: 1.11, 1.2, 1.33. The ratio of 1.2 is close to the "optimal" ratio found in the literature (Gaskill and Brown, 1990; Harris *et al.*, 1989).

II. METHODS

A. Stimuli and apparatus

The experimental setup for the measurement of the fine structure was similar to that used in the previous studies (He and Schmiedt, 1993, 1994). The primary tones (f_1 and f_2 , $f_2 > f_1$) were generated by two separate channels of a 16×16 bit waveform synthesizer (Pragmatic, 2201A-3X). The primaries were fed to two insert earphones (Etymotic Research, ER2) and delivered separately to the ear canal through two tubes. The tubes were housed in a low-noise microphone probe (Etymotic Research, ER10B). The earphone/microphone assembly was fitted to the ear canal by an impedance tip and the tubes protruded past the tip end by 7 mm, instead of 5 mm as recommended by the manufacturer (Etymotic Research). This modified tubing length was found optimal to prevent the standing-wave problem in the ear canal calibration (Siegel, 1994).

The sound-pressure level in the ear canal was recorded by the microphone, amplified (40 dB), and led to a spectrum analyzer (Hewlett Packard, 3561A). A 100-Hz frequency span was centered at the specific $2f_1 - f_2$ frequency and provided a frequency resolution of 0.25 Hz. The final spectrum was an average of four epochs. Each fine structure was obtained by measuring the ADP over a 1/3-octave frequency

range with a frequency interval of 1/32 octave.

Real-ear calibrations were performed using the ER10B microphone, and the resulting frequency responses were used to correct the earphone output to obtain a constant sound-pressure level across frequency.

B. Procedures

Six normal-hearing subjects (three male and three female) with ages between 20 and 37 years participated in this study. Each subject was tested in fourteen sessions, each session lasting about 2 h. The first two sessions repeated a previous study (He and Schmiedt, 1993), in which ADP fine structures were measured with equal-level primaries, incremented from 45 to 65 dB SPL in 2.5-dB steps. The measurements were made in 1/3-octave intervals centered at 2000 and 4000 Hz. In the remaining sessions, ADP fine structure was measured under two conditions: In condition 1, L_1 was fixed at 50 dB SPL, while L_2 varied from 30 to 75 dB SPL in 5-dB steps; in condition 2, the L_2 was fixed and L_1 varied over the same range. For each condition, ADP fine structure was measured with three primary frequency ratios (f_2/f_1): 1.11, 1.2, and 1.33. This yielded 12 sessions per subject (2 level ratios × 3 frequency ratios × 2 frequency regions).

In each session, SOAEs were monitored at the beginning and end of the measurements. Three subjects (JA, AT, and CR) had measurable SOAEs and these subjects had larger ADP levels in general. However, in only one case was an SOAE frequency close to a $2f_1 - f_2$ frequency (within 10 Hz) such that a significant interaction between the two components was observed during the measurement ("frequency locking," subject AT, at 2000 Hz, with $f_2/f_1 = 1.2$). Because the relationships between the ADP and SOAEs are beyond the scope of this study, these data are not discussed.

III. RESULTS

A. Downward frequency shift in the equal-level ($L_1 = L_2$) condition

Figure 2 shows examples of ADP fine structure from one subject measured with primary levels increasing from 45 to 65 dB SPL in 2.5-dB steps ($L_1 = L_2$; $f_2/f_1 = 1.2$). All functions except those obtained at the lowest level (45 dB) are sequentially offset by 5 dB for clarity. As level increases, the fine structure maintains its sharpness in the 2000-Hz range, but gradually smooths out in the 4000-Hz range. Furthermore, there is a tendency for the peaks and valleys to shift downward in frequency with the increasing level. For this subject, the downward shift was about 1/32–2/32 octave over a 20-dB range of primary level. These results are fully consistent with our previous study (He and Schmiedt, 1993).

B. Opposing frequency shifts with unequal primary levels

Figure 3 shows data from two subjects (top and bottom rows, respectively) measured under the two level-ratio conditions. The frequency ratio (f_2/f_1) was kept constant at 1.2. The ADP functions other than the ones taken at the lowest

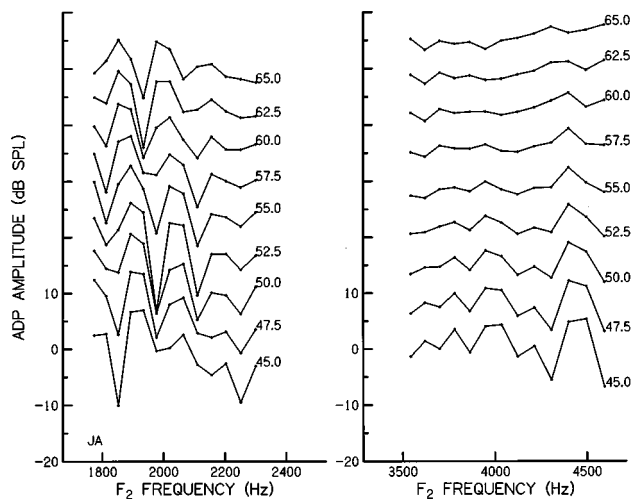


FIG. 2. ADP fine structures obtained from one subject measured at different primary levels near 2000 Hz (left) and 4000 Hz (right). The parameters are the levels for both primaries in dB SPL ($L_1 = L_2$). All functions except the lowest ones at 45 dB are sequentially shifted by 5 dB for clarity. The data obtained around 2000 Hz show a fine-structure shift toward lower frequencies with increasing level. At 4000 Hz, the fine structure becomes smoother with level and shows less shift. These data confirm the results of He and Schmiedt (1993).

level are sequentially shifted upward by 10 dB for clarity. Thus, the dB scale on the ordinate refers only to the lowest trace.

As shown in the first column of Fig. 3, with L_1 fixed at 50 dB SPL (condition 1), the peaks and valleys of the fine structure gradually shift upward in frequency as L_2 is increased from 30 to 75 dB SPL. On the other hand, a downward frequency shift is observed in the fine-structure pattern measured under condition 2 (second column), when L_2 is fixed and L_1 is increased. These shifts are consistent with our model predictions (Fig. 1). Although there was some inter-subject variance in the amount of shift, all subjects showed a tendency for the upward frequency shift (condition 1) to be somewhat less than the downward shift (condition 2). The two subjects in Fig. 3 showed a $1/32$ – $2/32$ octave upward shift, but a $3/32$ – $4/32$ octave downward frequency shift over a 40- or 45-dB range of the primary levels. The rate of the downward frequency shift with L_1 is similar to that observed with the equal-level primaries (Fig. 2).

The third column compares ADP fine structures obtained in separate test sessions when data were measured under either condition 1 or 2 for the case when both primary levels were equal to 50 dB SPL. The test/retest repeatability was excellent as indicated by the consistent alignment of the peaks and valleys, the high values of the correlation coefficients, and the small rms level differences between the two measures.

Figure 4 plots the fine structures obtained from two subjects measured around 4000 Hz. The data show a consistent tendency for a frequency shift with level like the data in Fig. 3. For condition 1, there is a clear upward frequency shift as a function of L_2 , although with a somewhat smaller rate. In

condition 2, for $L_2 > L_1$, the ADP fine structures tend to be smoothed and saturated in amplitude, obscuring the shift as compared to the data at 2000 Hz. However, for the functions where $L_2 \leq L_1$, a downward frequency shift is still clear. Thus, the amount of smoothing depends on which primary level was fixed. When L_1 is fixed at 50 dB SPL, the sharpness of the fine structure is basically unchanged as L_2 increases. On the other hand, when L_2 is fixed, the fine structure shows a gradual smoothing at high levels of L_1 . It seems that the smoothing or saturation of the fine structure is related to high levels of L_1 , but not L_2 , around 4000 Hz.

C. Effect of primary frequency ratio (f_2/f_1)

Figure 5 presents data similar to that in Fig. 3, but obtained with a frequency ratio of 1.11. As compared to the data with f_2/f_1 of 1.2, the patterns of the fine structures change with the smaller frequency ratio, even though the f_2 frequencies remain the same, suggesting changes in the ADP source. Moreover, the most striking difference between these two sets of data is in the rate of frequency shift. Although both sets show consistent trends of opposing frequency shifts, data with a ratio of 1.11 exhibit a significantly greater rate of frequency shift than those with ratio of 1.2. Over a 40-dB range of L_2 , the upward shift shown in Fig. 5 for condition 1 is $2/32$ – $4/32$ octave, a rate that is about twice that in Fig. 3 where $f_2/f_1 = 1.2$. An even more rapid rate of shift is evident in the data from condition 2. As shown in the second panel in the bottom row of Fig. 5 (subject KR), the right-most peak shifts about $9/32$ octave with a 35-dB increase in L_1 . Note that this shift with L_1 in condition 2 is not as gradual as that for $f_2/f_1 = 1.2$ (Fig. 3). Data from both subjects in Fig. 5 show a gradual shift up to 50 or 55 dB of L_1 , with an abrupt increase in the rate of shift as L_1 further increased. This abrupt shift was common for all subjects and was present only when $f_2/f_1 = 1.11$.

At 4000 Hz, three subjects (RT, CS, and JA) had very low ADP amplitudes with a f_2/f_1 ratio of 1.11, and noise-floor contamination made it difficult to assess the frequency shifts. But whenever sizable responses were obtained, they followed the a trend similar to that at 2000 Hz.

When the two primaries were placed farther apart in frequency ($f_2/f_1 = 1.33$), ADP amplitudes decreased substantially. Only five subjects showed measurable ADPs and among them only two subjects (AT and CR) had complete data sets measured at both the 2000- and 4000-Hz intervals. Even for those two sets, noise-floor contamination sometimes yielded random patterns of frequency shift with level. A complete set of data from one subject (AT) is shown in Fig. 6. Data measured around 2000 Hz are plotted in the top row; data measured around 4000 Hz are plotted in the bottom row. Substantially less frequency shift is evident as compared to the other frequency ratios in Figs. 3–5. Under condition 1 (first column, Fig. 6), the peaks and valleys basically remain at the same frequency locations as L_2 is increased. However, the ADP peaks tend to be skewed upward in frequency. Under condition 2 (second column), some downward frequency shift is just noticeable.

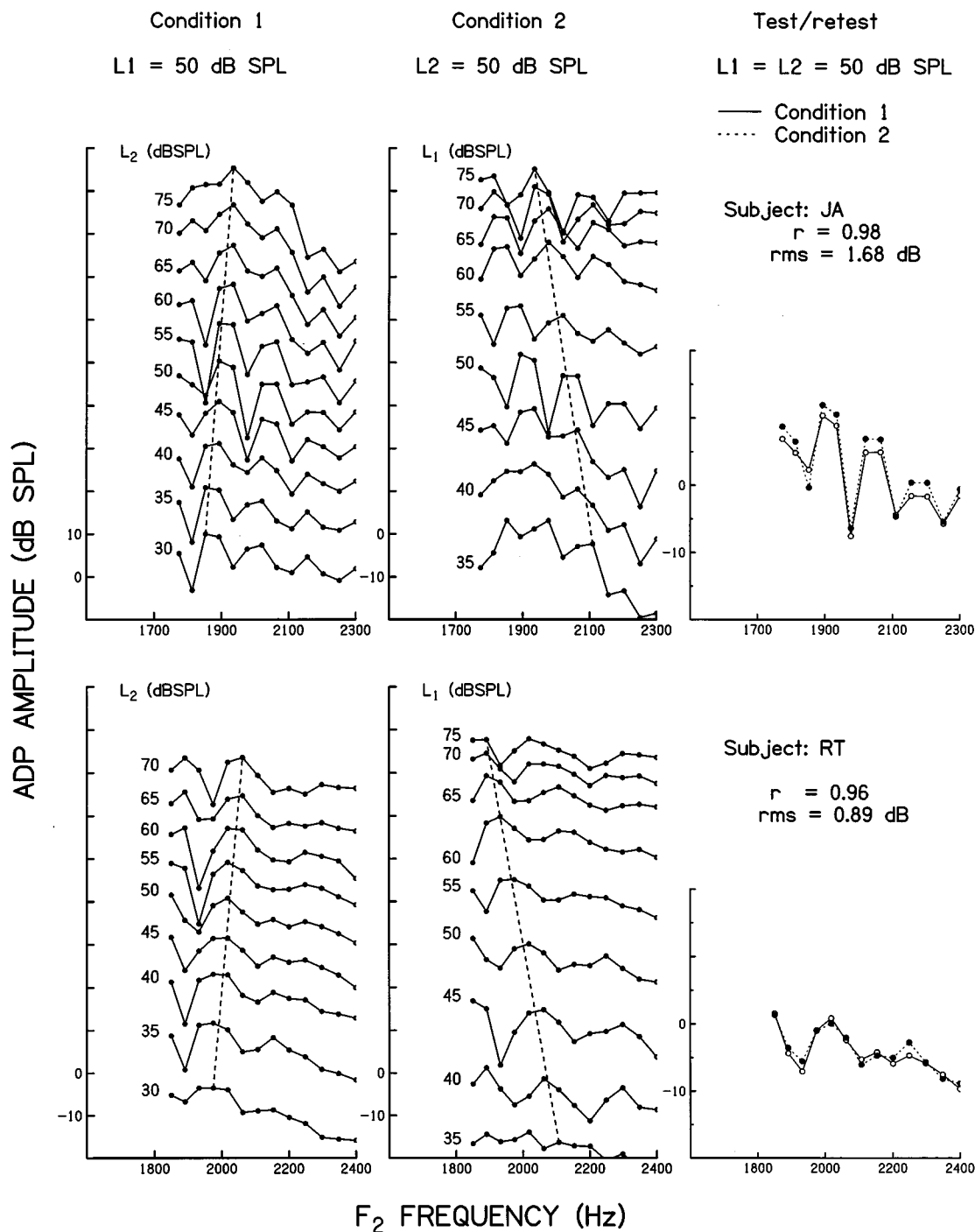


FIG. 3. ADP fine structures obtained from two subjects (upper and lower rows) measured under conditions 1 (first column) and condition 2 (second column) at frequencies around 2000 Hz. The primary frequency ratio was 1.2, and the level for the varying primary is indicated for each trace. Note the opposite frequency shifts of the ADP peaks and valleys in condition 1 and 2, as indicated by the dashed lines. The third column shows repeatability by plotting ADP fine structures obtained in separate sessions under conditions 1 and 2 in the case where $L_1 = L_2 = 50$ dB SPL.

D. Input/output functions

The f_2/f_1 ratio also affects the input/output (I/O) functions of the ADP amplitude, especially the maxima of the functions. Figure 7 shows examples of I/O functions measured from subject KR under conditions 1 (top) and condition 2 (bottom). Different f_2/f_1 ratios are represented by the columns. In each panel, the parameter is frequency starting from the lowest to the highest used in measuring the fine

structure around 2000 Hz. The scale on the ordinate refers to the I/O function obtained at the lowest frequency (1850.7 Hz); the remaining functions are sequentially offset by 5 dB for clarity (see legend). Typically, the I/O functions under condition 1 (top) have maxima when L_2 is less than 50 dB and have shallow slopes on both the increasing and decreasing sides. Under condition 2, the maxima generally occur when L_1 is greater than 50 dB SPL, and the slopes are

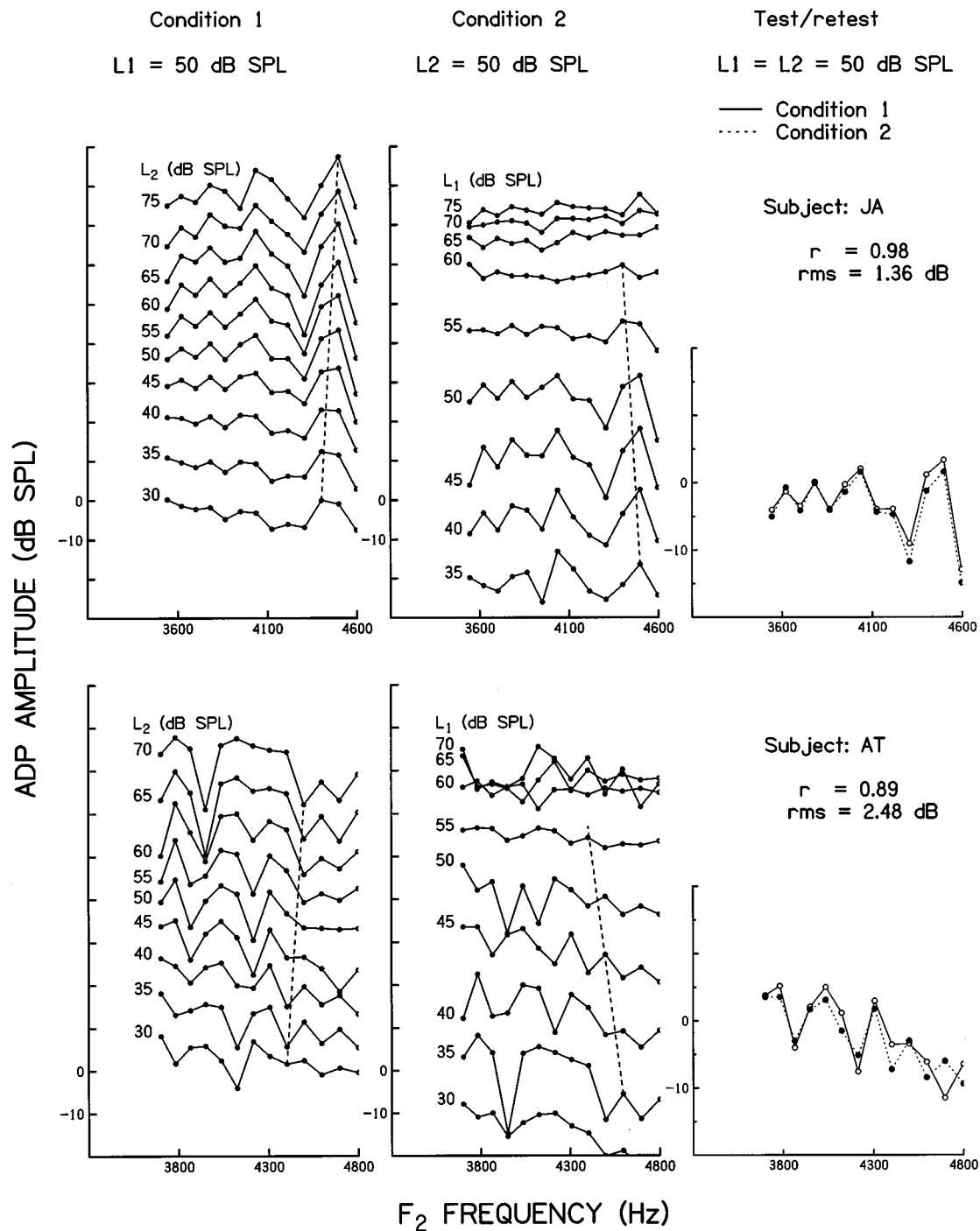


FIG. 4. Same as Fig. 4 but measured around 4000 Hz. Note that the fine structures become smoother with increasing levels of L_1 but not L_2 .

steeper than those in condition 1, especially with largest frequency ratio. These general tendencies are similar to those observed in a previous study of human and animal emissions (Brown and Gaskill, 1990).

Comparing data in different columns in Fig. 7, there is a clear effect of the f_2/f_1 ratio. Under condition 1, the slope of the I/O function tends to decrease with increasing frequency ratio. With $f_2/f_1 = 1.33$, the majority of the subjects showed flattened I/O functions, as compared to the smaller frequency ratios. Under condition 2, the slopes seem not to be affected by the f_2/f_1 ratio, but the I/O maxima shift to higher L_1

levels as the f_2/f_1 ratio increases from 1.11 to 1.33.

It is also clear that the shapes of I/O functions obtained at closely spaced frequencies within the fine structure can be quite different and may include the presence of nonmonotonic notches. In a previous study (He and Schmiedt, 1993), we demonstrated that these notches are directly related to the frequency shift of the ADP fine structure with increasing level. The data here further demonstrate that greater frequency shifts in the fine structure are associated with greater variance in the shapes of the I/O functions. For a frequency ratio of 1.11 (first column), more than one peak can be ob-

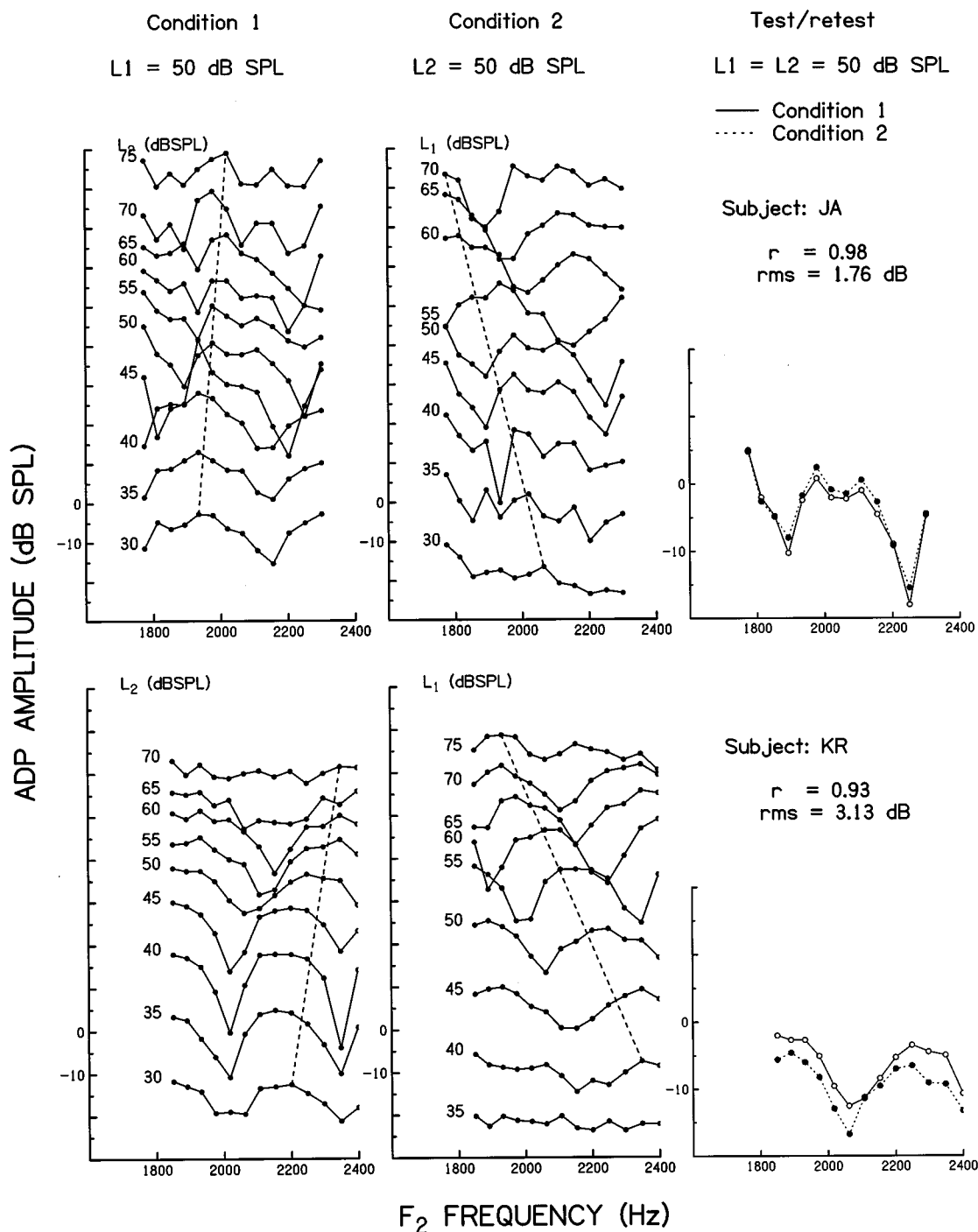


FIG. 5. Same as Fig. 4 but obtained with a f_2/f_1 ratio of 1.11. Note the much larger rates of shift as compared to those in Fig. 4.

served in many I/O functions, and this is almost certainly related to the rapid frequency shift of ADP fine structure with this frequency ratio (Fig. 5).

The I/O maxima which defines optimal L_1/L_2 ratio also varies, depending on the various conditions and parameters. Some general tendencies can be assessed from the overall distributions. Histograms of the optimal L_1 and L_2 levels obtained from all the subjects are shown in Fig. 8 for conditions 1 (left column) and 2 (right column). For both conditions, data centered at 2000 Hz are plotted in the top block; those centered at 4000 Hz are plotted in the bottom block. Within each block, the three panels present data for the three

f_2/f_1 ratios. The optimal L_1/L_2 ratio can be calculated knowing that either L_1 or L_2 is fixed at 50 dB SPL under the two respective conditions.

As can be seen from Fig. 8, distributions of the I/O maxima obtained under condition 1 are somewhat broader than those obtained from condition 2, indicating a larger intersubject variance. The larger variance is consistent with the observations of Brown and Gaskill (1990). Both data sets show generally well-defined peaks in the distribution except when $f_2/f_1 = 1.33$; here distributions are generally flat with less clearly definable peaks. These flat distributions are likely to be related to the reduced ADP amplitudes obtained with

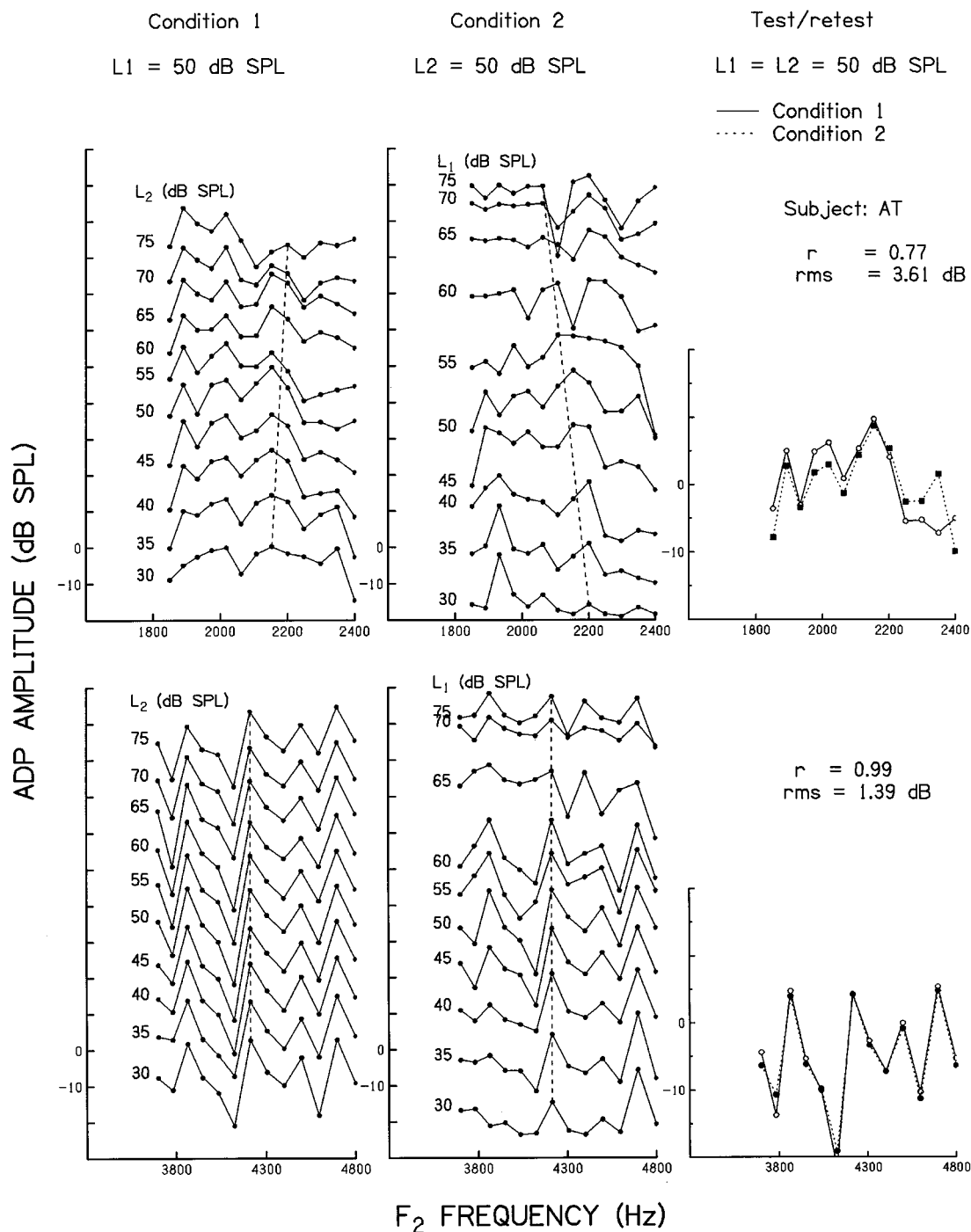


FIG. 6. Same as Fig. 4 but obtained with a f_2/f_1 ratio of 1.33 at frequencies around 2000 Hz (top) and 4000 Hz (bottom) from one subject. Little frequency shift with level is apparent with this frequency ratio.

this ratio as discussed above or, perhaps, to smaller sample sizes. Generally speaking, however, the distributions with $f_2/f_1 = 1.33$ show trends that are consistent with data for the other two frequency ratios.

Data in all the blocks in Fig. 8 show a consistent effect of f_2/f_1 ratio. Specifically, under condition 1 (left column), the I/O maxima tend to occur at higher L_2 levels as f_2/f_1 increase. A clear trend is seen around 4000 Hz under condition 1 (bottom block): The peak of the distribution is at 35 dB for L_2 when $f_2/f_1 = 1.11$, 40 dB when $f_2/f_1 = 1.2$, and 45 dB when $f_2/f_1 = 1.33$. In the 2000-Hz region under con-

dition 1 (top block), these trends are clearly seen in the data when $f_2/f_1 = 1.11$ and 1.2, and the distribution peaks at $L_2 = 40$ and 45 dB, respectively. For $f_2/f_1 = 1.33$, there is no clear peak, but the mean value of the distribution is at $L_2 = 46.1$ dB. Under condition 2 (right column), the peak of the distribution moves progressively to higher levels of L_1 as the f_2/f_1 ratio increases. These data indicate that the I/O maximum occurs at lower L_1 or L_2 at smaller f_2/f_1 ratios.

Finally, it is clear that primary frequency also affects the I/O maxima. Comparing data centered at 2000 Hz and 4000 Hz in Fig. 8 (top and bottom blocks), the optimal L_1/L_2

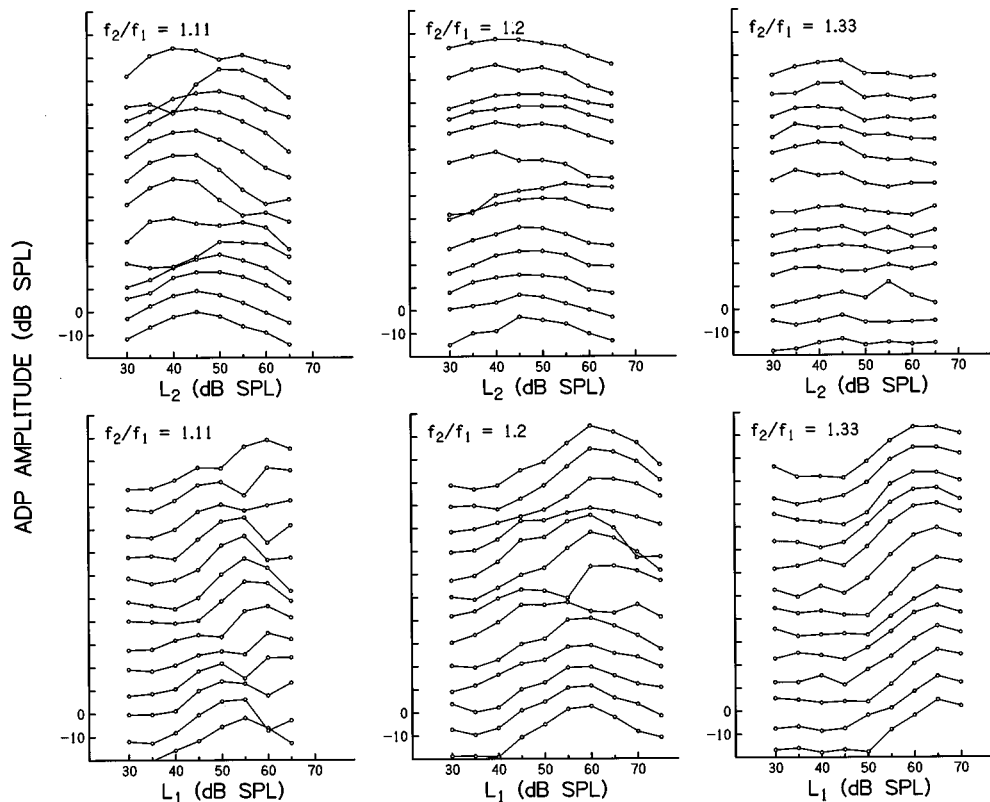


FIG. 7. Input/output (I/O) functions of the ADP obtained under conditions 1 (top) and 2 (bottom) with three different f_2/f_1 ratios (columns). In each panel, the I/O functions were measured at individual frequencies of the ADP fine structures centered at 2000 Hz. The f_2 frequencies for individual I/O functions are (from bottom to top) 1850.7, 1891.2, 1932.6, 1974.9, 2018.2, 2062.3, 2107.5, 2153.6, 2200.8, 2249, 2298.2, 2348.6, and 2400 Hz. The traces, except that for 1850.7 Hz, are sequentially offset by 5 dB for clarity, and the ordinate refers only to the lowest trace. When L_1 is fixed (top row), I/O functions are generally flat as compared to those when L_2 is fixed (bottom). Note that the I/O maxima tend to occur at higher L_1 or L_2 levels as the f_2/f_1 ratio increases from 1.11 to 1.33.

ratios for the 4000-Hz data are consistently 5 dB greater than their counterparts at 2000 Hz.

IV. DISCUSSION

A. Level-dependent frequency shift

In a previous study (He and Schmiedt, 1993), we observed a downward frequency shift of the ADP fine structure with increasing primary level ($L_1 = L_2$). The shift has been modeled as a result of the compressive nonlinearity present in the basilar-membrane response to increasing stimulus levels (Sun *et al.*, 1994a). This shift of the fine structure with primary level is confirmed and further extended by the data presented here, wherein one primary is varied in level while the other primary level is fixed. Under these conditions frequency shifts of the fine structure are in opposing directions as shown in Figs. 3–6. Specifically, when L_1 is fixed, the fine structure shows an upward shift in frequency with increasing L_2 (condition 1), whereas when L_2 is fixed and L_1 is increased, a downward shift is observed (condition 2). These results are consistent with the predictions of our model (Fig. 1) and suggest that the ADP fine structure largely reflects the place characteristics present in the overlapping area of the primary traveling waves.

An orderly change in the amount of the shift as a function of f_2/f_1 ratio is also observed. Significantly more fre-

quency shift is related to primaries with smaller frequency ratios (Fig. 5, $f_2/f_1 = 1.11$), as compared to when the two primaries are further apart in frequency (Figs. 3 and 4 for $f_2/f_1 = 1.2$ and Fig. 6 for $f_2/f_1 = 1.33$). These results support the hypothesis that the frequency shift is a function of amount of overlap. The amount of overlap can be related to the overall stimulation within the area of overlap, which in turn determines the nonlinear compression of the basilar-membrane response within this area.

There are concurrent changes in the maxima of the I/O functions as a function of f_2/f_1 (Fig. 7). The maxima occur at lower levels for the f_2/f_1 ratio of 1.11 than those for 1.2 and 1.33, suggesting more compression or “suppression” by one primary on the other. This tendency is more clearly shown in Fig. 8, where the distribution of the maxima of the I/O functions tends to progressively move up in primary level as f_2/f_1 ratio changes from 1.11 to 1.33, indicating less compression when the primary frequencies are farther apart. This trend is consistent with the frequency shift of the fine structure: Increased amounts of frequency shift are associated with increased amounts of compression.

Another finding is that L_1 is more effective in causing a frequency shift in the fine structure than L_2 . As shown in Figs. 3–6, the downward frequency shift (condition 2 with changing L_1) is generally at a higher rate than the upward frequency shift with (condition 1 changing L_2). This can also

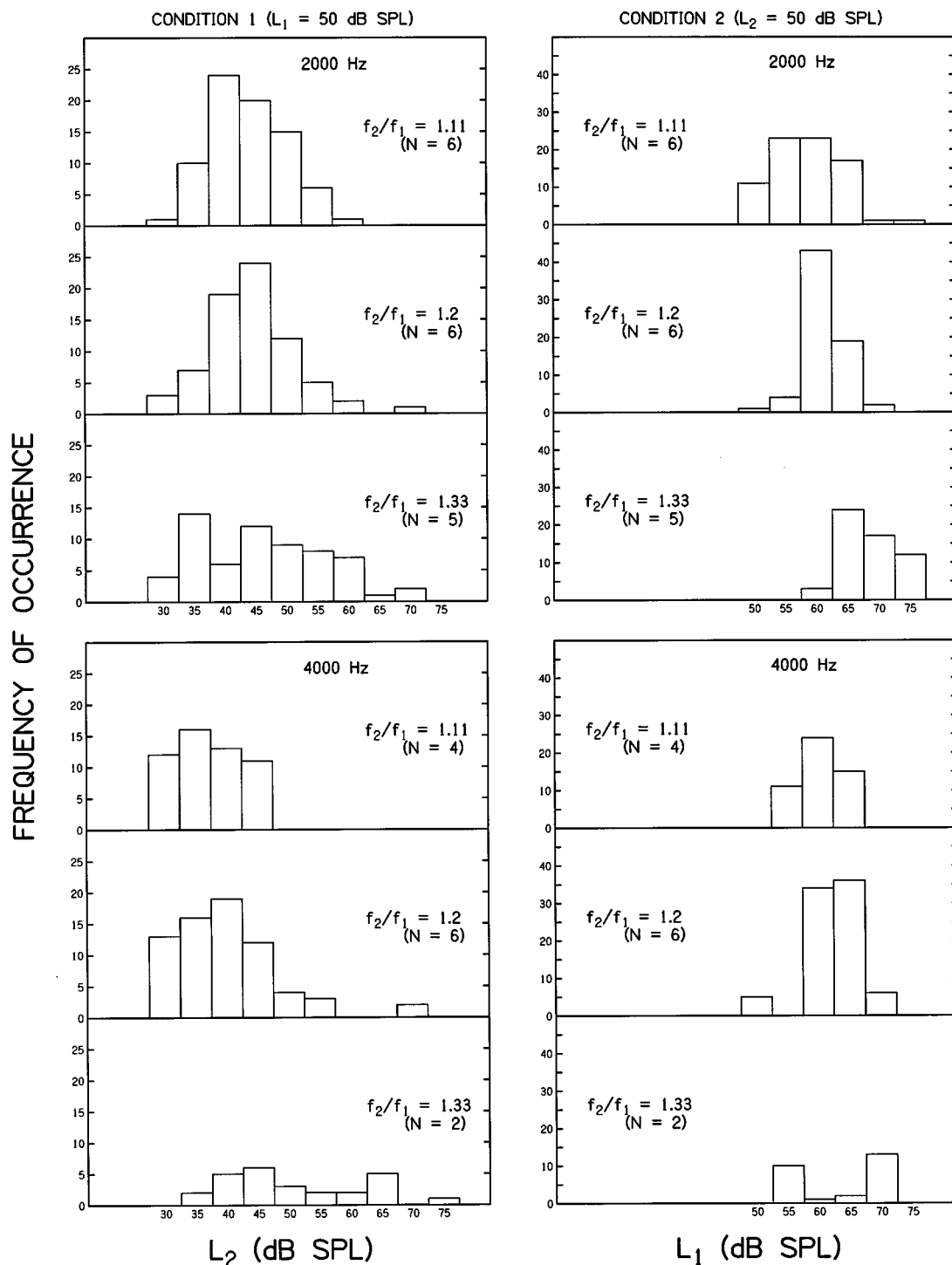


FIG. 8. Distributions of the optimal L_2 and L_1 for conditions 1 (left column) and 2 (right column), respectively, summarized from all the subjects (see text). The two blocks within each column represent data obtained around 2000 Hz (top) and 4000 Hz (bottom). Within each block, data are segregated with respect to the frequency ratio of the primaries. Note that the peaks of the distributions tend to occur at higher levels of L_2 or L_1 as the f_2/f_1 ratio increases.

be explained in terms of the overall stimulation within the overlap area under the two conditions (see Fig. 1). As originally demonstrated by Rhode (1971) and more recently by Ruggero and Rich (1991), the excitation pattern of the basilar membrane becomes broader (expanding at both sides) with increasing stimulus level. At the same time, the peak moves basalward. Under condition 1 where L_1 is fixed and L_2 is varied, the frequency shift is mainly determined by the

expansion of the f_2 traveling wave (on the apical side), and the basalward shift of the peak seems not to increase the amount of overlap or overall stimulation within the overlap area. In condition 2 when L_1 increases, both expansion of the f_1 traveling wave (on the basal side) and basalward shift of the peak excitation contribute to the overall stimulation in the area of overlap, resulting in more frequency shift than in condition 1. A similar explanation can be used in the situa-

tion when both primary levels are increased concurrently ($L_1 = L_2$). Thus, the rate of the downward frequency shift observed in condition 2 corresponds to that in the equal-level situation.

Similar frequency shifts have been observed in the suppression of the SFE (Brass and Kemp, 1993) and the ADP (Kummer *et al.*, 1995). In spite of the differences in the stimulation for these two types of emissions, both sets of data show that the lower-frequency suppression tones become more effective than higher-frequency suppressors as the suppressor level or suppression criterion increase. The key factor here is the compressive nonlinearity and the frequency shift of basilar-membrane tuning at increasing stimulus levels. The excitation peak for a given frequency shifts basalward with higher levels (Rhode, 1971; Ruggero and Rich, 1991). Therefore, with increasing suppressor levels, the excitation patterns for both the high- and low-frequency suppressors shift toward the basal end, but the high-frequency suppressors move away from the emission region in the overlap area, whereas the low-frequency suppressors shift toward the region. Hence, for a given primary frequency ratio, high-frequency tones have less effect on the fine structure with level than low-frequency tones.

B. Irregularity in the damping factor for the ADP generator

One of the important findings of the modeling study by Sun *et al.* (1994a) is that a statistical irregularity in the damping factor is a necessary condition for the model to generate the fine structure. Without this specific irregularity, essentially no rippling pattern can be simulated, even with the rapid phase rotation in the resonant regions of the primary traveling waves (Zwicker and Peisl, 1990). Sun *et al.* examined other alternatives as sources of the irregularity: (1) nonuniform phase alone; (2) nonuniform amplitude alone; and (3) nonuniform amplitude and phase. It was found that none of these variables when varied stochastically could generate any sizable fine structure. They suggested that the irregularity in the amplitude and phase of the traveling wave must be coordinated, governed by the damping factor, in order to produce the fine structure. The importance of an irregularity was also recognized in later modeling studies of the spectral periodicity in the SFE (Zweig and Shera, 1995) and the ADP (Talmadge *et al.*, 1996). Like our model, Talmadge *et al.* also noticed that without a stochastic variable, the ADP function was basically flat, but with only a small amount of place-fixed “micromechanical irregularity” a dramatic ripping pattern could be generated.

A conceptual difference between our model and the model of Talmadge *et al.* (1996) is that the latter model only considers the irregularity which gives rise to the “reflections” to be localized in the $2f_1 - f_2$ region. The model of Sun *et al.* seems more realistic in assuming that the irregularity exists throughout the basilar membrane. The $2f_1 - f_2$ wavelets generated within the overlap area in the primary region have dominant contributions (larger amplitudes) to the resultant ADP fine structure, as compared to the re-emission components from the $2f_1 - f_2$ place.

In reality, there may indeed be re-emissions involved in the generation of ADP fine structure. There is evidence that the $2f_1 - f_2$ component generated at the primary place travels to its characteristic place (Kim *et al.*, 1980). Since this component acts like an external tone that causes a traveling-wave pattern at the $2f_1 - f_2$ place, an SFE is likely to be obtained. What the relative contribution of the SFE from the $2f_1 - f_2$ place is to the ADP measured in the ear canal is the important question. According to the results of this study, as interpreted with the model of Sun *et al.* (1994a, b), the contribution of the SFE is small as compared to that from primary place. It is difficult to explain in terms of re-emissions the observed frequency shifts in opposing directions under the two conditions, as well as the effect of the primary frequency ratio on the rate of shift. Rather, a more parsimonious explanation is to assume that all the level-dependent frequency shifts of the ADP fine structure arise largely from the nonlinear response of the basilar membrane within the overlap area of the primary traveling waves.

V. SUMMARY

- (1) When L_1 is fixed and L_2 is varied, an upward frequency shift is observed in the ADP fine structure; when L_2 is fixed and L_1 is varied, a downward shift in frequency is observed. These opposing frequency shifts are consistent with predictions from the model of Sun *et al.* (1994a,b), suggesting that the ADP fine structure largely reflects the place features within the overlapping area of the primary traveling waves.
- (2) The downward frequency shift with increasing L_1 shows a significantly larger rate of shift than the upward shift with increasing L_2 .
- (3) An orderly difference in the rate of frequency shift as a function of the f_2/f_1 ratio is observed, with the greatest rate of shift associated with the smallest f_2/f_1 ratio.
- (4) These results suggest that the frequency shift is a function of the amount of overlap of the primary traveling waves or the overall stimulation in the area of overlap. The underlying mechanism for the frequency shift can be modeled as arising from the compressive nonlinearity of the basilar membrane in response to increasing stimulus levels.
- (5) The optimal L_1 or L_2 for the generation of the ADP is affected by the f_2/f_1 ratio, such that the I/O maxima occurs at lower primary levels at smaller ratios of f_2/f_1 .

ACKNOWLEDGMENTS

We thank Drs. F. A. Boettcher, J. R. Dubno, and two anonymous reviewers for their critiques of earlier manuscripts. This work was supported by NIH/NIDCD Grants Nos. RO3DC-02674 and PO 1 DC-00422.

- Brass, D., and Kemp, D. T. (1993). “Suppression of stimulus frequency otoacoustic emissions,” *J. Acoust. Soc. Am.* **93**, 920–939.
- Brown, A. M., and Gaskill, S. A. (1990). “Measurement of acoustic distortion reveals underlying similarities between human and rodent mechanical responses,” *J. Acoust. Soc. Am.* **88**, 840–849.

- Brown, A. M., and Kemp, D. T. (1984). "Suppressibility of the $2f_1-f_2$ stimulated acoustic emission in gerbil and man," *Hear. Res.* **13**, 29–37.
- Dolan, T. G., and Abbas, P. J. (1985a). "Changes in $2f_1-f_2$ acoustic emission and whole-nerve response following a sound exposure: Long-term effects," *J. Acoust. Soc. Am.* **77**, 1475–1483.
- Dolan, T. G., and Abbas, P. J. (1985b). "Short-term effects of sound exposure on $2f_1-f_2$ acoustic emission," *J. Acoust. Soc. Am.* **77**, 1614–1616.
- Furst, M., Rabinowitz, W. M., and Zurek, P. M. (1988). "Ear canal acoustic distortion at $2f_1-f_2$ from human ears: Relation to other emissions and perceived combination tones," *J. Acoust. Soc. Am.* **84**, 215–221.
- Gaskill, S. A., and Brown, A. M. (1990). "The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity," *J. Acoust. Soc. Am.* **88**, 821–839.
- Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (1989). "Acoustic distortion products in humans: Systematic changes in amplitude as a function of f_2/f_1 ratio," *J. Acoust. Soc. Am.* **85**, 220–229.
- Harris, F. P., Probst, R., and Xu, L. (1992). "Suppression of $2f_1-f_2$ otoacoustic emission in humans," *Hear. Res.* **64**, 133–141.
- He, N., and Schmiedt, R. A. (1993). "Fine structure of the $2f_1-f_2$ acoustic distortion product: Changes with primary level," *J. Acoust. Soc. Am.* **94**, 2659–2669.
- He, N., and Schmiedt, R. A. (1994). "Effects of aging on the fine structure of the $2f_1-f_2$ acoustic distortion product," *J. Acoust. Soc. Am.* **99**, 1002–1015.
- He, N., Dubno, J. R., and Schmiedt, R. A. (1994). "Frequency fine structures of auditory thresholds and distortion product emissions: Is there a correlation?" *Abst. Assoc. Res. Otolaryngol.* **17**, 48.
- Kemp, D. T., and Brown, A. M. (1983). "An integrated view of cochlear mechanism nonlinearities observed from the ear canal," in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever (Delft U.P., Delft), pp. 75–82.
- Kemp, D. T., and Brown, A. M. (1984). "Ear canal acoustic and round window electrical correlates of $2f_1-f_2$ distortion generated in the cochlea," *Hear. Res.* **13**, 39–46.
- Kim, D. O., Molnar, C. E., and Matthews, J. W. (1980). "Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure," *J. Acoust. Soc. Am.* **67**, 1704–1721.
- Kimberley, B. P., Shaw, G., and Shera, C. A. (1995). "Cochlear acoustic reflectance and traveling wave delay," *J. Acoust. Soc. Am.* **97**, 3413.
- Kummer, P., Jassen, T., and Arnold, W. (1995). "Suppression tuning characteristics of the $2f_1-f_2$ distortion-product otoacoustic emission in humans," *J. Acoust. Soc. Am.* **98**, 197–210.
- Martin, G. K., Probst, R., Scheinin, S. A., Coats, A. C., and Lonsbury-Martin, B. L. (1987). "Acoustic distortion products in rabbits. II. Sites or origin revealed by suppression and pure-tone exposures," *Hear. Res.* **28**, 191–208.
- Piskorski, P., Long, G. R., Talmadge, C. L., and Tubis, A. (1995). "Origin of the fine structure of distortion product emissions in the human ear," *Abst. Assoc. Res. Otolaryngol.* **18**, 119.
- 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.
- Ruggero, M. A., and Rich, N. C. (1991). "Application of a commercially manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration," *Hear. Res.* **51**, 215–230.
- Schmiedt, R. A. (1984). "Acoustic injury and physiology of hearing: A review and tutorial," *J. Acoust. Soc. Am.* **76**, 1293–1317.
- Schmiedt, R. A. (1986). "Acoustic distortion in the ear canal. I. Cubic difference tones: Effect of acute noise injury," *J. Acoust. Soc. Am.* **79**, 1481–1490.
- Siegel, J. H. (1994). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," *J. Acoust. Soc. Am.* **95**, 2589–2597.
- Schloth, E. (1983). "Relation between spectral composition of spontaneous otoacoustic emissions and fine-structure of threshold in quiet," *Acustica* **53**, 250–256.
- Sun, X., Schmiedt, R. A., He, N., and Lam, C. F. (1994a). "Modeling the fine structure of $2f_1-f_2$ acoustic distortion product (ADP) I. Model development," *J. Acoust. Soc. Am.* **96**, 2166–2174.
- Sun, X., Schmiedt, R. A., He, N., and Lam, C. F. (1994b). "Modeling the fine structure of $2f_1-f_2$ acoustic distortion product (ADP) II. Model evaluation," *J. Acoust. Soc. Am.* **96**, 2175–2183.
- Talmadge, C. L., Piskorski, P., Tubis, A., and Long, G. R. (1996). "Evidence for multiple spacial origins of the fine structure of distortion product otoacoustic emissions in humans, and its implications—Experimental and modeling results," *Abst. Assoc. Res. Otolaryngol.* **19**, 94.
- Wilson, J. P. (1980). "The combination tone, $2f_1-f_2$, in psychophysics and in ear-canal recording," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. Van den Brink and F. A. Bilsen (Delft U.P., Delft), pp. 43–50.
- Wit, H. P., Langevoort, J. C., and Ritsma, R. J. (1981). "Frequency spectra of cochlear acoustic emissions ('Kemp echoes')," *J. Acoust. Soc. Am.* **70**, 437–445.
- Zweig, G., and Shera, C. A. (1995). "The origins of periodicity in the spectrum of evoked otoacoustic emissions," *J. Acoust. Soc. Am.* **98**, 2018–2047.
- Zwicker, E. (1990). "On the frequency separation of simultaneously evoked otoacoustic emissions' consecutive extrema and its relation to cochlear traveling wave," *J. Acoust. Soc. Am.* **88**, 1639–1641.
- Zwicker, E., and Peisl, W. (1990). "Cochlear preprocessing in analog models, in digital models and in human inner ear," *Hear. Res.* **44**, 209–216.
- Zwicker, E., and Schloth, E. (1984). "Interrelation of different oto-acoustic emissions," *J. Acoust. Soc. Am.* **75**, 1148–1154.

Are normal hearing thresholds a sufficient condition for click-evoked otoacoustic emissions?

Sarosh Kapadia^{a)}

MRC Institute of Hearing Research, Ropewalk House, 113 The Ropewalk, Nottingham NG1 6HA,
United Kingdom

Mark E. Lutman

Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ,
United Kingdom

(Received 6 June 1996; revised 3 February 1997; accepted 25 February 1997)

Transiently evoked otoacoustic emissions (TEOAE) have been reported in several studies as absent in a small minority of normal ears. Other studies have reported TEOAEs in all normal ears. Differences between studies may arise directly from criteria for TEOAE identification, criteria for selection of normals, or statistically due to limited sample sizes. In order to understand and model cochlear processes involved in TEOAE generation, it needs to be known whether the presence of normal hearing leads automatically to generation of TEOAEs. The present study set out to establish in a large sample if any ears could be found that lacked TEOAEs despite normal hearing threshold levels (HTL). A total of 397 ears from highly cooperative adult subjects were examined under laboratory conditions. Using cross correlation between replicate nonlinear waveforms as the criterion, TEOAEs were present in 99.2% of the sample (lower CI 98.1%). However, careful visual assessment of the recorded waveforms for the remaining ears did not unequivocally show absence of TEOAE characteristics in any ear with normal HTLs. While TEOAE strength varies widely among ears, no clear evidence was found to show that TEOAEs can be absent when HTLs are normal. © 1997 Acoustical Society of America. [S0001-4966(97)05906-7]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Bt [RDF]

INTRODUCTION

Transiently evoked otoacoustic emissions (TEOAE) represent the release of acoustic energy from the cochlea into the ear canal (Kemp, 1978). The existence of a TEOAE is held to indicate normal or near-normal outer hair-cell function, and it is generally accepted that the presence of TEOAEs implies hearing threshold levels (HTL) are 30 dB or better at least at some frequencies in the range 0.5–4 kHz (Probst *et al.*, 1991), barring rare cases of purely retrocochlear disorder. However, it is unclear whether the converse is always true, as a number of studies have been unable to detect TEOAEs in a small minority of adult ears with normal hearing (Rutten, 1980; Probst *et al.*, 1986; van Dijk and Wit, 1987; Stevens, 1988; Dolhen *et al.*, 1991; Lutman and Saunders, 1992; Prieve *et al.*, 1993).¹

Failure to record a TEOAE in a normally hearing ear may be due to limitations of the measuring conditions (recording system, ambient conditions, scoring criteria), or alternatively TEOAEs could be genuinely absent in some normal ears. The latter would imply that normal cochlear (and middle-ear) function is not a *sufficient* condition for the generation of TEOAEs. It is also well established that in many normal ears TEOAE energy may be measurable at some frequencies of normal hearing, but not at adjacent frequencies also having normal hearing (Probst *et al.*, 1991). This again suggests that factors other than those required for normal

hearing can influence the generation of TEOAEs, although destructive interference between TEOAE components may offer an alternative explanation. This paper represents the first part of an investigation into the reasons why some normal adult ears display absent or weak TEOAEs, thereby gathering insights into the generation and fundamental properties of TEOAEs.

The literature is generally consistent in showing a prevalence above 95% for TEOAEs in ears with normal hearing,² excluding studies where recording conditions were not well controlled. Some studies have obtained TEOAEs in all normal ears tested; notably several studies by Bonfils and co-workers, of which Bonfils *et al.* (1988a) and Bonfils *et al.* (1988b) are representative, those of Vedantam and Musiek (1991) and Reshef *et al.* (1993), being based on 131, 105, 100, and 61 normal ears, respectively. Other studies such as Kemp (1978) have also reported emissions in all normal ears tested, but in small samples. Four studies have reported prevalence figures of 96%–97% (Probst *et al.*, 1986; Stevens, 1988; Dolhen *et al.*, 1991; Lutman and Saunders, 1992).

Part of this variation among studies may arise from statistical uncertainty in estimates obtained from small samples. Clearly, there is always the possibility that TEOAEs could be obtained in all ears in a random sample, even if the true population prevalence is less than 100%. To minimize this possibility, larger samples are required than those quoted above.³ However, if a genuine sample prevalence of anything less than 100% is obtained, then the population prevalence logically *cannot* be 100% (barring system limitations).

^{a)}Now at MRC Institute of Hearing Research, Southampton Outstation, Royal South Hants Hospital, Southampton SO14 0YG, UK.

Aside from statistical uncertainties, three further issues are important in considering studies of TEOAE prevalence among normal ears, particularly those reporting 100% prevalence. These concern the criteria for defining normal ears and for identifying TEOAEs.

First, most prevalence studies of normals have excluded subjects with ear-related irregularities, even if they have normal HTLs. These exclusion criteria have included history of ear disease, exposure to noise, exposure to ototoxic agents, upper respiratory tract infection at the time of testing, and minor otoscopic aberrations. Bonfils *et al.* (1988b) also required stapedius reflex thresholds to be normal. On the other hand, Vedantam and Musiek (1991) merely required that subjects “were without otologic or hearing complaints” and that middle-ear compliance was 0.4 mmho or greater, and Reshef *et al.* (1993) only required that their subjects were free of “middle-ear problems.” While it is not established whether these factors influence TEOAEs *without* affecting thresholds, it is possible that prevalence figures are dependent on selection criteria additional to HTL.

Second, authors have differed in their criteria for normal HTL, and group thresholds may have varied substantially among studies. Stevens (1988) specified an upper limit based on mean HTL (0.5, 1, 2, and 4 kHz) of 20 dB, whereas Bonfils *et al.* (1988b) required HTLs for all frequencies (0.25, 0.5, 1, 2, 4, and 8 kHz) to be 10 dB or better. This strict criterion also applied to 70% of the normal ears in Bonfils *et al.* (1988a), representing subjects below the age of 40 years. (For subjects between 40 and 50 years and between 50 and 60 years, who also all had emissions, thresholds are reported as being normal relative to the subject’s age, but are only reported as means for the two groups.)

Third, the criteria for deciding whether an emission is present or absent may vary between studies. Most of the above studies appear to have made this decision primarily on the basis of subjective rating by experienced scorers. Such scorers presumably take a combination of factors into account, and can more readily allow for contaminating noise and discount subtle artifacts that can influence most objective scoring schemes (Stevens, 1988; Lutman, 1993; Salomon *et al.*, 1993a). Even when objective measures such as cross correlation between a pair of replicate TEOAE waveforms are used, lack of standardization between different probes, recording equipment, and methods of calculation makes direct comparisons among studies questionable. The situation is further complicated by differences between (a) those measurement techniques that allow *only* the derived “nonlinear” component of the TEOAE to be recorded and analyzed, and (b) those that record and store a set of raw (often termed “linear”) emission waveforms from which a nonlinear component can be derived in addition to the linear components.⁴

In reviewing the available literature, and notwithstanding the reports of failure to measure TEOAEs in some normal ears, Probst *et al.* (1991) state “it is likely that all normal ears have transiently evoked otoacoustic emissions when tested carefully with specialized laboratory equipment.”⁵ They suggest that TEOAEs are a general property of the normal human peripheral auditory system and point to tech-

nical reasons as being responsible for reports of “absent” TEOAEs in normal ears. This proposition is scientifically parsimonious and accounts for (a) the weight of evidence that prevalence in normal ears is very close to 100%, and (b) the theoretical understanding of the commonality of the biomechanical processes which underlie normal cochlear sensitivity and the generation of TEOAEs. Given the statistical impossibility of establishing that all normal ears have TEOAEs, it is reasonable to adopt this proposition, in the absence of clear evidence to the contrary. However, if it were shown convincingly that TEOAEs can be absent in some ears under good recording conditions, despite normal HTLs, the proposition would fail. Such a finding would be important to the understanding and modeling of cochlear active and mechanical processes. If normal cochlear sensitivity can occur without TEOAE generation, then there must be at least a partial dissociation of the processes leading to normal sensitivity and to the generation of TEOAEs.

The aim of the present work was to search for ears with normal HTLs but no evidence of a TEOAE. To increase the probability of finding such ears, a large sample was tested under well-specified laboratory conditions. In so doing, the study would also establish a reliable estimate for the prevalence of TEOAEs, which would have value for setting the upper limit of specificity that could be achieved by hearing screening programs based on TEOAEs. As a by-product, subjects without TEOAEs, but with normal HTLs, would be identified for further detailed study into the reasons for their absent TEOAEs. If no such subjects were identified, ears with weak TEOAE would be identified for similar purposes.

I. METHODS

A. Subjects

Potential subjects were drawn from two main sources—one group from research, clinical and support staff working at the authors’ institute and associated clinical departments, and the second from students at the attached university. Subjects in the second group were paid for participating in the study. This approach avoided the potentially biased sample that would occur by testing patients attending an audiology clinic, but found to have normal HTLs. Written informed consent was obtained from all subjects in the study.

As the main aim of the study was to search for ears with both normal HTLs and absent TEOAEs, ears could be excluded from the search as soon as it was discovered that they had either TEOAEs or abnormal hearing. In view of the large numbers to be tested, the following strategy was adopted for the sake of efficiency. Ears were excluded from the search at any of three stages, arranged in order of ease of implementation: self-reported hearing difficulty, then TEOAE screening, and finally follow-up test sessions which included detailed TEOAE testing and audiometry.

Self-reported hearing status was ascertained initially at an interview where subjects were asked if they had any hearing difficulties, such as in using the telephone or in other day-to-day activities. All ears for which no such difficulties were reported were classed as “presumed normals” and went on to be screened for TEOAEs. Any ears that gave

TABLE I. Age distributions for female and male subjects in the basic test group.

| Age band (years) | Male | | Female | |
|------------------|----------|----|----------|----|
| | <i>N</i> | % | <i>N</i> | % |
| 10–20 | 19 | 22 | 21 | 18 |
| 20–30 | 45 | 53 | 59 | 51 |
| 30–40 | 13 | 15 | 25 | 22 |
| 40–50 | 6 | 7 | 6 | 5 |
| 50–60 | 2 | 2 | 5 | 4 |

weak or absent emissions were then tested by audiometry and those with HTLs outside the criteria given below were eliminated from the search. In addition, a subgroup of ears with normal emissions also underwent audiometry for control purposes (see Sec. III). This contingent approach obviated audiometry in most presumed normals and afforded significant economies of test time. However, it meant it was uncertain whether all subjects with TEOAEs had normal HTLs. This was of no consequence to the main proposition examined by the study, but entails that our prevalence estimate based on “presumed normals” rather than verified normals contains a degree of uncertainty. Section III addresses this issue and demonstrates that the degree of uncertainty is negligible, as was obvious from the outset.

In all, 203 subjects attended the initial interviews. Two subjects reported unilateral hearing loss verified by prior audiological testing and these two ears were excluded. Thus 404 “presumed normal” ears from 203 subjects were screened for TEOAEs. Of these, 7 ears of 5 subjects were found at audiometry to have HTLs outside the study criteria and were hence eliminated from the analysis, leaving 397 ears from 201 subjects. These 397 ears will be referred to as the basic test group of the study. Of the basic test group, 230 ears (58%) were from female subjects. Subject age ranged from 18.2 to 59.6 years. Table I shows the age distributions of the 201 subjects, for males and females separately.

B. Audiometric criteria

Audiometric thresholds were determined at 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz. Normal hearing was defined by air-conduction HTLs no greater than 20 dB at any test frequency between 0.25 and 4 kHz. These frequencies were chosen because TEOAEs do not tend to have frequency components above about 5 kHz (Probst *et al.*, 1991)—indeed most recording systems, including the one used here, have bandwidths limited to about 5 kHz.

C. Equipment

TEOAEs were measured using the Programmable Otoacoustic Emission Measurement System (POEMS) described by Cope and Lutman (1988), which was run in its “automatic” mode. This equipment comprises a probe containing a miniature earphone and microphone, a click generation system, preamplification, and signal conditioning circuitry,

all controlled by a personal computer with time-domain averaging software. It has been widely used in research and clinical practice in the U.K.

Unipolar (condensation) clicks, obtained by applying rectangular pulses with a duration of 100 μ s to the probe earphone, were presented at a rate of 40/s. Click levels used for each test were 70, 60, and 50 dB peak-equivalent (pe) SPL, as measured at the reference microphone in a “2-cc” coupler conforming to IEC 126. The measurement system incorporates an overload rejection facility, which eliminates raw time records that contain any sound-pressure amplitudes which exceed preset limits, as can occur due to subject movement or breathing noises. The remaining “good” time records are then averaged. The overload rejection limits were set for each individual test, at a level which would reject about 5% of all raw time records. This statistical scheme is an effective way to improve signal-noise ratio for a given number of averages, and was preferred to an arbitrary fixed criterion equal for all subjects.

One thousand “good” time records were interleaved into two separate averaging buffers at each click level. The time records corresponded to the period from 4 to 20 ms following the click. The click waveform generated in each ear was also recorded by separately averaging two replications of 100 time records each, starting at electrical click onset, for a click level of 50 dB pe SPL.

Where audiometry was performed, a clinical audiometer (Grason-Stadler GSI 16 fitted with TDH-50 earphones) calibrated according to ISO 389 was used.

D. Procedure

All TEOAE measurements were performed by the first author, who is experienced in the measurement of emissions. Soft plastic eartips were used to secure the measurement probe in the ear canal. Close attention was paid to the selection of tip size, the placement of the probe, and the routing of the probe lead such that it was not in contact with the subject’s body or chair. Subjects were seated comfortably inside a sound-isolated booth, with the tester and the main parts of the emission recording system outside the booth. Subjects were asked to avoid swallowing and to remain as still and relaxed as possible while measurements were made. At the conclusion of the recording in each ear, the position of the probe in the subject’s ear canal was checked. In the event of any probe slippage apparent to the tester or reported by the subject, or if any significant subject movement during the measurement had occurred, a repeat recording was performed after refitting the probe and reinstructing the subject. In all such instances both sets of recordings were analyzed, and the clearer recorded emission was taken as the result for that ear for that test session.

II. RESULTS

A. Preliminary analysis

The nonlinear or saturating component of the TEOAE was computed by extracting a pair of “nonlinear” waveforms from the pairs of replicate “linear” waveforms measured at 70 dB and 60 dB pe SPL. The original linear re-

sponses at 70, 60, and 50 dB pe SPL were preserved and available for further analysis.⁶ This contrasts with the more common practice of only preserving the nonlinear component.

Cross correlation coefficients between replicate linear waveforms measured at 70, 60, and 50 dB pe SPL, and also between the derived nonlinear pair, were computed for the 6- to 16-ms segments. This time window avoids potential stimulus artifacts that can occur at the start of the waveform, and potentially poor signal-noise ratio at the end.

Historically, the identification of TEOAEs has relied heavily on expert visual evaluation of the measured waveforms. Although such approaches may be less prone to obvious errors than are simple objective means of identification (Stevens, 1988; Lutman, 1993; Salomon *et al.*, 1993a), visual scorers may differ where marginal judgements are to be made. Thus objective means of identification are important for comparisons between different studies. We have based our prevalence estimate on the cross correlation between replicate nonlinear waveforms (r_{nl}). Nonetheless, we recognize its potential shortcomings and have also utilized expert visual rating.

The cross correlation measure r_{nl} has been widely used in the analysis of TEOAEs (e.g., Kemp *et al.*, 1986; Stevens, 1988; Harris and Probst, 1991; Priewe *et al.*, 1993). Its use is also supported by our previous work that attempted to determine reliable objective scoring methods to be used instead of visual rating (Lutman, 1993). Of several measures used singly and in combination, none was a more reliable predictor of visual scoring than r_{nl} (using the 6–16 ms time segment). In the present study the calculation of the correlation coefficient was preceded by off-line high-pass filtering of the measured waveforms, in order to reduce the influence of low-frequency noise. Tognola *et al.* (1995) have confirmed the value of *post hoc* high-pass filtering as an aid to TEOAE identification. Waveforms yielding values of r_{nl} below 0.5 were classified as “fails” (emissions absent or questionable) and those with r_{nl} greater than or equal to 0.5 as “passes” (emissions present). A criterion value of approximately 0.5 has been reported by others (e.g., Kemp *et al.*, 1986; Stevens, 1988), and is currently widely used in many centers (Dirckx *et al.*, 1996), although differences between test systems and choice of waveform time segments mean that exact equivalence is not assured. We did not attempt to discriminate between “absent” and “questionable” TEOAEs based on an arbitrary value of r_{nl} , none having received general acceptance in the literature.

All visual rating was performed independently by the second author (who has similarly rated several thousand such waveforms in various studies). Sets of waveforms from the “screening” test sessions were classified simply as “passes” (i.e., TEOAEs present) or “fails” (TEOAEs absent or questionable), taking all original and derived nonlinear replicate pairs into account. Waveforms recorded in the repeat test sessions were rated in more detail using a slightly modified version of the five-point subjective rating system described previously by Lutman (1989). For the present study, only three rating categories were used: “absent,” “questionable,” and “present.” These correspond to the rat-

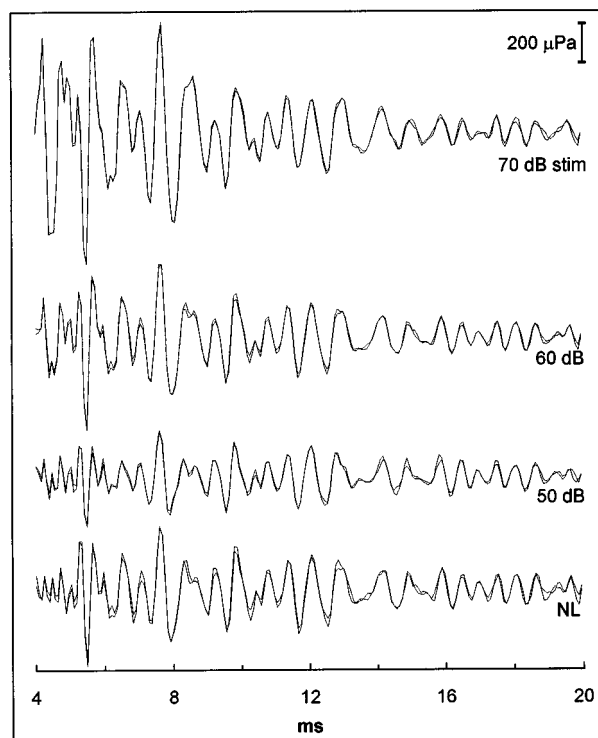


FIG. 1. TEOAE waveform pairs from an ear exhibiting clear emissions. Four pairs of superimposed waveforms are shown; replicate waveform pairs for click levels of 70, 60, and 50 dB pe SPL, and the nonlinear waveform pair derived from the waveform pairs for click levels of 70 and 60 dB. The horizontal axis indicates elapsed time following delivery of the click. The waveforms exhibit several features commonly seen in TEOAEs, which are described in the text. The cross correlation coefficient between the nonlinear waveforms, calculated for the time segment 6 to 16 ms (r_{nl}) is 0.98.

ing points “0,” “1,” and “2 through 4,” respectively, described with illustrative examples in Lutman (1989).

To illustrate the above analyses and to provide a framework for discussions to follow, Fig. 1 shows a typical set of clear TEOAEs measured in the present study. The figure shows the measured replicate responses to clicks at 70, 60, and 50 dB pe SPL and the replicate nonlinear responses [labeled NL] derived from the responses to the clicks at 70 and 60 dB pe SPL. The nonlinear waveforms are calculated such that linearly scaling components (in the corresponding measured responses) cancel, while the amplitudes of fully saturated components are exactly preserved. Partially saturated components are reduced in amplitude, but do not cancel.

Many of the features commonly exhibited by TEOAEs are seen in the waveform set of Fig. 1.

- (i) The replicate waveforms are highly correlated.
- (ii) A recognizably similar waveform pattern occurs in all four replicate pairs.
- (iii) A degree of frequency dispersion occurs, such that higher-frequency components tend to occur earlier and lower-frequency components later.
- (iv) Comparing traces labeled 70 and 60 dB, the waveform amplitudes are almost equal from the middle through to the end of the time records; that is, the emission is largely saturated. This is reflected by the amplitude in this part of the waveform being much the

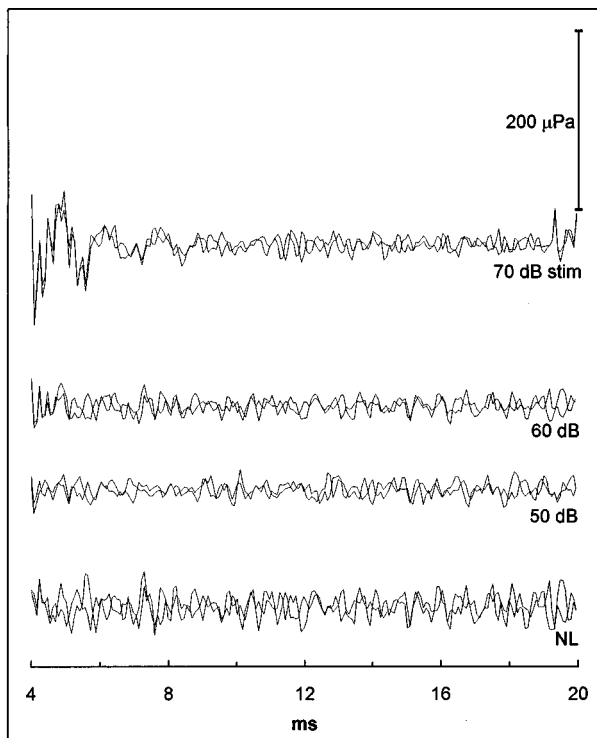


FIG. 2. TEOAE waveform pairs from a hard-walled 0.5 ml cavity. The figure layout is as for Fig. 1. The cross correlation coefficient between nonlinear waveforms (r_{nl}) is 0.07.

same in the nonlinear response. On the other hand, the amplitude at the start of the waveform is clearly smaller in the 60-dB trace than in the 70-dB trace, as is reflected in the substantially reduced amplitude in the corresponding part of the nonlinear waveform.

Our experience of conducting these measurements suggests that the period comprising the first 1–2 ms shown of each waveform is often dominated by the tail of the click stimulus, rather than representing a true TEOAE. However, in the example of Fig. 1, notice that total cancellation of this earliest part of the waveform does not occur in NL. The residual high-frequency component is possibly a short-latency TEOAE, but may be a small artifact arising from system nonlinearities. This component is well correlated between the replicate waveforms, and its existence illustrates why excluding the first 6 ms of the time record is helpful when calculating r_{nl} . The r_{nl} value obtained here is 0.98.

Figure 2 shows the waveforms obtained exactly as above, but with the probe inserted into a hard-walled cavity with a volume of 0.5 ml. The waveforms show little evidence of correlated signals, other than a small, very-low-frequency component in the early part of the top trace, upon which is superimposed a high-frequency signal similar to the artifact referred to in Fig. 1. The waveforms would be rated “absent” for TEOAEs; the r_{nl} value obtained is 0.07. Similar recordings are obtained in ears with severe hearing impairment, albeit with greater noise having physiological origin.

B. Cross correlation coefficients between nonlinear waveforms

Examination of the 397 ears of the basic test group and application of the cross correlation criterion to the initial “screening” test measurements resulted in 11 ears classified as “fails” ($r_{nl} < 0.5$). Of these 11 ears, 10 were tested in a second test session and the waveforms obtained again classified on the basis of r_{nl} . Of the ten sets of waveforms from repeat test sessions, three yielded r_{nl} values less than 0.5, and were thus finally classified as “fails.” The remaining seven ears passed at the second test session, which was conducted under less time pressure and in a different test booth, which provided better isolation at low frequencies. Prevalence based on r_{nl} is therefore 393 out of 396 ears, or 99.2%. To calculate this prevalence figure we have excluded the one ear that gave an r_{nl} value less than 0.5 upon initial testing, but did not undergo a repeat test session and did not have HTLs checked.

C. Results of visual rating

Of the 397 ears in the basic test group, 9 were rated as “fails” (TEOAEs absent or questionable), based on visual rating of the initial screening measurements. Eight of these nine were in the group similarly classified on the basis of r_{nl} above. (The remaining 388 that were rated as passes included the ear that failed initially on r_{nl} but was not retested.) All nine “fails” were tested in a second test session and the waveforms rated again. Of the nine second sets of waveforms, *none* received a rating of “absent,” and only two a rating of “questionable.” Emissions in the remaining seven ears were rated as “present.” Hence, the overall outcome of our evaluation of the 397 ears based on visual rating of both linear and nonlinear TEOAE waveforms is: TEOAEs “present” in 395 ears, “questionable” in two ears, and “absent” in zero ears. We were thus unable to identify a single ear in our sample in which the TEOAE waveform set showed no evidence of an emission, having excluded ears with abnormal HTLs.

Table II lists the values of r_{nl} and the visual ratings from both TEOAE test sessions, for each of the 11 ears that underwent a repeat test session as a result of “fail” classifications from either scheme. The table demonstrates a satisfactorily high degree of correspondence between the two independent assessment schemes. The repeat session results are of main interest here, as they represent the final outcomes. It can be seen that the three ears finally classed as “fails” on r_{nl} (ears A, C, and D, $r_{nl} < 0.5$) include one of the two results visually rated as questionable (ear A), but not the other (ear B). The waveforms from ears C and D are visually rated as having TEOAEs present, despite having r_{nl} values less than 0.5.

D. TEOAE waveforms for marginal results

Figures 3 and 4 respectively, show the two TEOAE waveform sets that were visually rated as “questionable” (ears A and B). Neither exhibits the extent of repeatability between replicate waveforms seen in the example of Fig. 1. To the extent that repeatable TEOAE-type waveforms are

TABLE II. Visual ratings of waveform sets and cross correlation coefficients between replicate nonlinear waveforms (r_{nl}), for initial and repeat TEOAE measurements. Results are shown for each of the 11 ears that underwent a repeat test session as a result of “fail” classification from either scheme.

| Ear | Initial session | | Repeat session | |
|-----|---------------------|----------|----------------|----------|
| | Visual rating | r_{nl} | Visual rating | r_{nl} |
| A | absent/questionable | 0.12 | questionable | 0.25 |
| B | absent/questionable | 0.57 | questionable | 0.78 |
| C | absent/questionable | 0.16 | present | 0.40 |
| D | absent/questionable | 0.08 | present | 0.49 |
| E | absent/questionable | 0.20 | present | 0.51 |
| F | absent/questionable | 0.09 | present | 0.53 |
| G | absent/questionable | 0.26 | present | 0.54 |
| H | absent/questionable | 0.37 | present | 0.59 |
| I | absent/questionable | 0.28 | present | 0.68 |
| J | present | 0.49 | present | 0.81 |
| K | present | 0.12 | present | 0.86 |

present in Figs. 3 and 4, these are predominantly in the early parts of the “linear” waveforms recorded in response to the highest click levels and show a more or less uniform decay from the very start of the time records. A cautious interpretation could be that these components represent unusually long-lasting ringing of the clicks rather than an emission, and

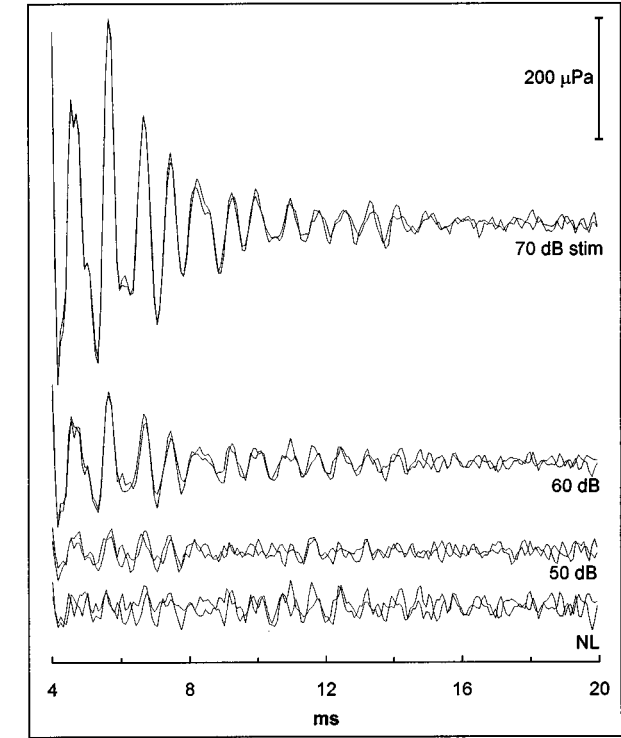


FIG. 3. TEOAE waveform pairs from repeat test session of ear A. The figure layout is as for Fig. 1. The cross correlation coefficient between nonlinear waveforms (r_{nl}) is 0.25.

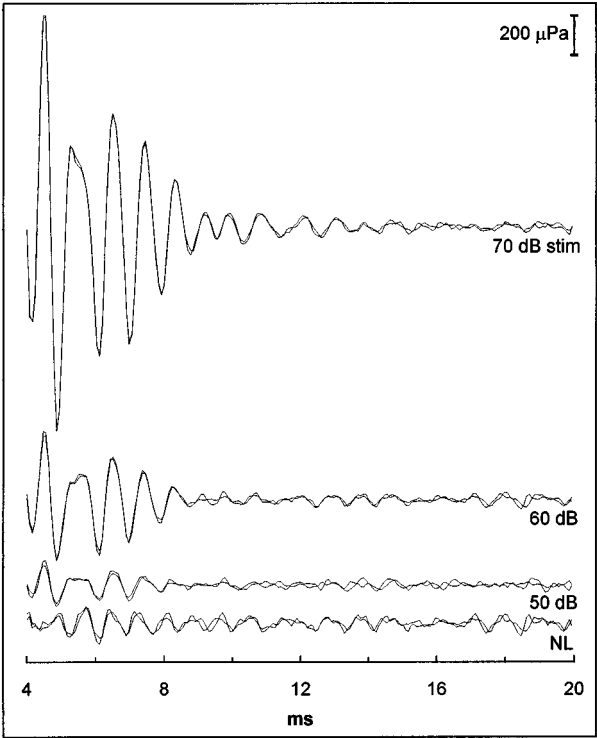


FIG. 4. TEOAE waveform pairs from repeat test session of ear B. The figure layout is as for Fig. 1. The cross correlation coefficient between nonlinear waveforms (r_{nl}) is 0.78.

TEOAEs were therefore rated as “questionable” rather than “present.” In the case of Fig. 4, although this early component is not completely cancelled in the nonlinear waveform (thus accounting for the high value of r_{nl}), it is also at a lower frequency than is typical for a genuine TEOAE. Finally, it should be stressed that the “questionable” ratings here were not associated with poor test conditions, such as excessive ambient or subject-generated noise. Neither of the waveform sets which were rated “questionable” demonstrated an unusually high level of noise compared with the norm for TEOAE waveforms recorded in this study. (Compare, e.g., Fig. 1 with Figs. 3 and 4.)

Figures 5 and 6, respectively, show the two TEOAE waveform sets from the repeat test sessions that yielded r_{nl} values less than 0.5, but which were visually rated as having emissions present (ears C and D). Although the emissions in these two ears are not as clear and repeatable as in the example of Fig. 1, in both cases a clear TEOAE, distinct from the tail of the stimulus, can be seen.

Figures 3–6 represent TEOAE waveforms which are at the margins of detectability and serve to contrast the results of visual rating and classification on r_{nl} . The reasons for the discrepant results arising from these two approaches are discussed further in the next section.

Table III gives the HTLs for ears A–D, obtained on the same days as the TEOAE waveforms shown in Figs. 3–6. They confirm that the four ears met our normal audiometric criterion. Furthermore, barring ear D (in which the failure on r_{nl} is most clearly erroneous), thresholds were also within normal limits at 6 and 8 kHz.

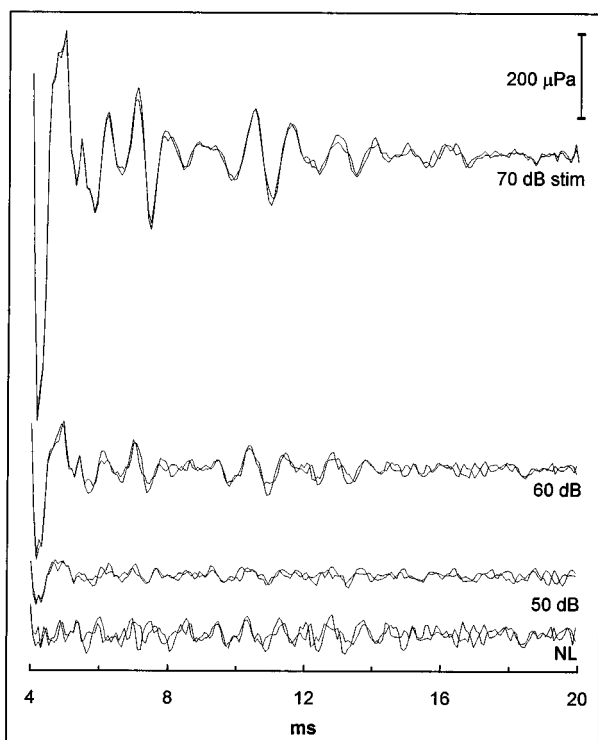


FIG. 5. TEOAE waveform pairs from repeat test session of ear C. The figure layout is as for Fig. 1. The cross correlation coefficient between non-linear waveforms (r_{nl}) is 0.40.

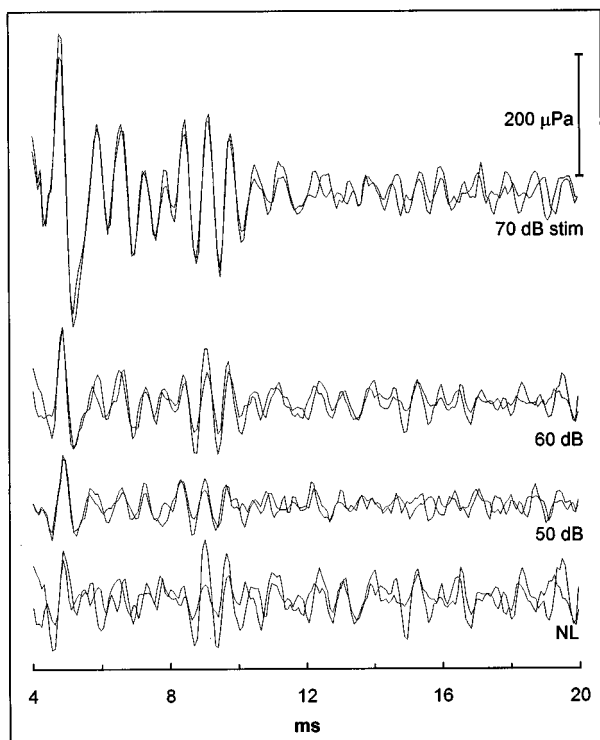


FIG. 6. TEOAE waveform pairs from repeat test session of ear D. The figure layout is as for Fig. 1. The cross correlation coefficient between nonlinear waveforms (r_{nl}) is 0.49.

TABLE III. Hearing threshold levels for ears A–D.

| Ear | Frequency (kHz) | | | | | | | |
|-----|-----------------|-----|----|----|----|-----|----|----|
| | 0.25 | 0.5 | 1 | 2 | 3 | 4 | 6 | 8 |
| A | 5 | 5 | 5 | 15 | 0 | 0 | 10 | 5 |
| B | 0 | 0 | 5 | 15 | 5 | –10 | 10 | 20 |
| C | 0 | –5 | –5 | 0 | 10 | 0 | 10 | 10 |
| D | –5 | 5 | 20 | 15 | 5 | 20 | 30 | 35 |

III. DISCUSSION

A. Prevalence of TEOAEs in normals

One aim of the study was to estimate the prevalence of TEOAEs in adult ears with normal hearing. Using the objective parameter r_{nl} to identify TEOAEs yields a prevalence estimate of 99.2%. The use of a cut-off r_{nl} value of 0.5 reflects a historical context of hearing screening, where there is a deliberate bias toward failing (i.e., regarding TEOAEs absent in) marginal cases. Further, being a relatively simple measure, r_{nl} is also prone to recognizable errors (some of which are discussed in Sec. III B below) in both sensitivity and specificity. Nonetheless it is a widely used basis for identifying TEOAEs, and we have adopted it for our prevalence estimates to facilitate comparisons across studies. However, the r_{nl} criterion of 0.5 is inherently conservative, which suggests we should also regard our prevalence estimate as conservative, possibly underestimating the true prevalence of TEOAEs in normal ears.

All “presumed normal” ears that were classified as “fails” on r_{nl} (as well as those rated “questionable” on visual rating) underwent audiometry. Therefore, such classifications in our study could not be explained by undetected hearing impairment. However, we did not perform audiometry on all ears that passed. It is a simple matter to demonstrate by sensitivity analysis that even gross violations of the presumption of normal hearing have only a minimal influence on the prevalence estimate. For that purpose we posit that 20% of the presumed normal ears that passed on r_{nl} classification would not have met our audiometric criterion.⁷ This would reduce the audiometrically normal sample to 318 ears. Three of these gave r_{nl} values below our TEOAE criterion, and hence the prevalence in the sample would be 315/318 instead of 393/396 (99.1% rather than 99.2%). Hence, the prevalence estimate is insensitive to our presumption of normal hearing in ears not tested audiometrically.

As the TEOAE prevalence figure based on r_{nl} is inherently conservative, and the value obtained is close to 100%, it is inappropriate to ascribe an upper confidence limit. Thus we have calculated a one-sided 95% confidence interval (CI), extending from 100% down to a prevalence of 98.1%.⁸ Given the large size of our sample and the resulting narrow confidence interval on our prevalence estimate, it is likely that previous studies that obtained significantly lower TEOAE prevalence figures were conducted using systems, protocols, or analyses that result in poorer TEOAE sensitivity. It is of course possible that future studies using more sensitive systems than ours might detect unambiguous

TEOAEs in 100% of a comparable number of ears. The present study simply indicates the lower bound of population prevalence of TEOAEs in normal adult human ears.

The two-stage testing protocol, with all ears that failed on r_{nl} classification from the initial screening sessions being tested a second time, afforded a degree of security in assessing marginal cases. In the majority of cases where measurements were repeated, there was a reduction in the noise in the waveforms recorded in the repeat sessions, leading directly to increased measures of r_{nl} . Subjects may have been better able to remain physically still during the second test sessions, the need for this having again been reinforced. The repeat tests were also conducted at a more leisurely pace over extended sessions. This emphasizes the need for careful laboratory measurement (coupled with the use of clearly defined criteria) in studies that seek to estimate the prevalence of TEOAEs.

B. Absence of TEOAEs

Cross correlation, r_{nl} , is a somewhat arbitrary measure for the identification of TEOAEs, but compares well with visual rating in most cases. However, in marginal cases the two methods can give discrepant results (Table II). Therefore, when attempting to establish the absence of a recordable TEOAE with certainty in exceptional cases, it is necessary to examine the limitations of both methods and to compare discrepancies. Part of the reason for discrepancies is the superior ability of an experienced scorer to reject certain kinds of artifact, and also to identify an emission in a noisy waveform. These two factors are illustrated in the TEOAE waveforms from ears B and D (Figs. 4 and 6) and are probably largely responsible for the discrepant outcomes in these two ears. A further important difference between the visual rating and the objective classification, as performed in this study, is the fact that the former makes use of all four waveform pairs, while the latter uses only the derived nonlinear waveform pair. For this reason it is instructive to consider what visual ratings may be given to the nonlinear waveform pairs alone in Figs. 5 and 6. These two waveform sets were visually rated as having TEOAEs present, but yielded r_{nl} values less than 0.5. Arguably, TEOAEs would be rated "absent" in both ears if based on the nonlinear waveforms only. Thus at least some of the discrepancy between the visual rating and the objective classification outcomes in this study is explicable in terms of the different information on which these two assessment methods were based, rather than on the greater sensitivity of expert visual rating *per se*. This reinforces our contention that the availability of linear TEOAE waveform data is important for the assessment of the emission, regardless of the choice of assessment method. The same point has been made by other workers (e.g., Salomon *et al.*, 1993b), in particular relation to measurement of TEOAEs in neonates.

A fundamental reason for a poorer rating (visual or objective) arising from evaluation of the nonlinear trace pairs alone in general is that TEOAEs in some subjects are substantially nonsaturated at the stimulus levels used to derive the nonlinear waveform. The less saturated a waveform component, the greater its degree of cancellation in the derivation

of the nonlinear waveform. Examination of Fig. 5 in particular supports this explanation in the case of ear C. Poorer ratings may also arise from the reduction in signal-noise ratio inherent in the derivation of the nonlinear waveform (Kemp *et al.*, 1990).

The preceding discussions raise the philosophical issue of what exactly should be construed as representing a TEOAE. Ideally, this would be according to the definition used in the Introduction, the release or return of acoustic energy from the cochlea into the ear canal, in response to an acoustic transient. In practice (and particularly in measurements in human ears), it is seldom possible to substantiate that a particular recorded signal is of cochlear origin.⁹ In evaluating recorded waveforms, therefore, an operational definition is required: for example, a repeatable signal that is distinguishable from the ringing of the stimulus click. Additional features of the response, such as amplitude saturation and frequency-latency relationships, provide secondary guidance, as discussed above. In our experience of thus evaluating TEOAE recordings, however, a gray area in between TEOAE "presence" and "absence" is inescapable, and we have been forced to admit a "questionable" category in visual rating of waveforms.

On this basis, accepting that TEOAEs are present in the recordings from ears C and D (Figs. 5 and 6) seems justified, despite the low r_{nl} values. In the case of ears A and B, the repeatable components evident in the waveforms (Figs. 3 and 4) cannot be dismissed unequivocally as stimulus artifact. In summary, we are therefore unable to identify unequivocally any ears in our sample with normal HTLs but no evidence of TEOAE.

C. Theoretical implications

It has been suggested that the generation of TEOAEs requires an inherent, perhaps anatomical, irregularity along the basilar membrane (e.g., Wilson, 1980; Sutton and Wilson, 1983; Kemp, 1986; Furst and Lapid, 1988; Wit *et al.*, 1994). This hypothesis implicitly allows for occasional findings of absent TEOAEs in normal ears, as rare occurrences of anatomically perfect cochleae might exist. Hypotheses that do not incorporate such irregularities require alternative means to allow for "anomalous" absences of TEOAEs if they exist.¹⁰ Such hypotheses may be simplified and made more tenable by the removal of this requirement, if all normal ears generate TEOAEs.

The frequency spectra of TEOAEs commonly exhibit unexplained gaps, in that little emission energy is measurable at some frequencies of normal hearing, despite demonstration of stimulus energy at the same frequencies. Two broad explanations could be postulated for this observation: (a) Some areas of a normal cochlea are unable to generate TEOAEs; or (b) All areas of a normal cochlea can (and perhaps do) generate TEOAEs, but interactions or other influences suppress emissions of some frequencies.¹¹ Alternative (a) is more consistent with the notion of TEOAEs being generated by anatomical irregularities and would also suggest that normal cochleae which do not generate TEOAEs might exist. To the extent that we were unable to demon-

strate clear examples of such ears, our findings would tend to support alternative (b).

We chose at the outset of the study not to apply any subject exclusions other than those based on air-conduction thresholds. This contrasts with some earlier studies, as noted previously, in which subjects were excluded due to factors such as past ear disease, exposure to noise or ototoxic agents, elevated stapedius reflex thresholds, presence of otoscopic aberrations, or upper respiratory tract infection at the time of testing. We indeed wanted such conditions to be admissible in our subject sample, for the very purpose of investigating their influence on TEOAE prevalence. The finding that virtually all of the 397 ears in our basic test group demonstrated TEOAEs shows that any such minor auditory factors have little effect on TEOAE prevalence, insofar as they may have been present.

We also ignored HTLs at and above 6 kHz in applying exclusions, in accordance with the broad consensus that TEOAEs provide no information on HTLs at frequencies other than those present in the emission spectrum. Thus HTLs at 6 kHz and above are not expected to have any direct association with TEOAE prevalence. Indeed, some authors (Rutten, 1980; Johnsen *et al.*, 1993) have concluded that HTLs up to only 2 kHz are of primary importance to TEOAE generation. Somewhat contrary to this view, Avan and co-workers (Avan *et al.*, 1991; Avan *et al.*, 1993; Avan *et al.*, 1995) have suggested that damage to the basal region only of the cochlea (as evidenced by increased high-frequency pure-tone thresholds) can reduce the amplitudes and alter the waveform patterns of TEOAEs in both guinea pigs and humans.¹² Additionally, Hauser *et al.* (1991) compared TEOAEs in ears with high-frequency sensorineural hearing loss with those from normal ears, by analyzing amplitudes of the TEOAEs across emission frequency. They found lower TEOAE amplitudes in the ears with hearing loss, even at the frequencies of normal hearing. Were mild-to-moderate high-frequency hearing loss (corresponding to basal cochlear pathology) to influence TEOAE presence in humans, we may have expected some absent emissions from the approximately 10% of our subject pool aged 40 to 60 years (Table I). Our failure to detect such ears suggests that hearing loss at high frequencies alone (to the degree that may be expected in subjects aged 40–60) does not materially affect TEOAE prevalence in humans. Not only did we fail to detect any ears with “absent” emissions that passed our restricted audiometric criterion, but the only two ears with “questionable” ratings had HTLs no greater than 20 dB at 6 and 8 kHz (Table III).

In summary, we have found no clear evidence for a dissociation between the cochlear processes leading to normal HTLs and to the generation of TEOAEs in response to clicks. It therefore seems likely that the condition of normal HTLs is indeed sufficient to ensure TEOAE generation. The question remains as to whether the low amplitudes of TEOAEs occurring in some normal ears represents an anomaly, or whether this can be explained on the basis of other measures of auditory function or status. Further investigation of the ears identified as having relatively weak TEOAEs and suitable controls will attempt to answer that

question, by gathering detailed measurements of other types of otoacoustic emissions in the same ears, and by conducting a range of other detailed measurements of auditory function. This should shed some further light on the general variation of otoacoustic emissions across ears and any association with other properties of hearing. Those studies may ultimately help to answer the question why markedly different emission levels can be recorded in ears with similar hearing threshold levels.

IV. CONCLUSIONS

(1) The prevalence of TEOAEs in normal adult ears, identified on the basis of the cross correlation between replicate nonlinear waveforms, is estimated to be 99.2%, with a one-sided 95% lower confidence limit of 98.1%. Within the range of HTL accepted into the present study, up to 20 dB at frequencies between 0.25 and 4 kHz, any minor auditory factors that may have been present can have little effect on TEOAE prevalence.

(2) This prevalence estimate is inherently conservative. Careful visual evaluation of the waveforms could not rule out TEOAEs in any ears.

(3) The amplitudes of TEOAEs, and hence the signal-noise ratios of recordings, vary considerably across ears with normal hearing. Less sensitive measurement systems, less ideal test conditions, or less cooperative subjects may yield prevalence figures somewhat lower than that reported here. The variation of TEOAE amplitude across audiometrically normal ears may arise from aspects of auditory function that are not evident in pure-tone air-conduction thresholds. Further research is required to explore this possibility.

(4) We were unable to identify any ear in our normal sample that unequivocally lacked TEOAEs, and are therefore unable to reject the scientifically parsimonious proposition that normal HTLs are a sufficient condition for TEOAE generation.

ACKNOWLEDGMENTS

We thank Pauline Smith and Melanie Ferguson for their advice and help over the duration of this study, and our subjects for their cooperation and patience. The manuscript was much improved by the comments of Mark Haggard and Alan Palmer, and the statistical advice of David Marshall and Adrian Davis. This work was supported in part by Defeating Deafness (formerly Hearing Research Trust), 330/332 Gray's Inn Road, London WC1X 8EE.

¹There have also been many reports of TEOAE prevalence in healthy neonates (e.g., Stevens *et al.*, 1987; Bonfils *et al.*, 1988c; Johnsen *et al.*, 1988; Kok *et al.*, 1993; White *et al.*, 1994), that have reported prevalences of the order of 95%–97%. However the major difficulties in ensuring good recording conditions and in accurately ascertaining HTLs mean prevalence reports of less than 100% are difficult to interpret.

²A notable exception is the study by van Dijk and Wit (1987) who reported the markedly low figure of 40%. It has been suggested by Probst *et al.* (1991) that this low figure was largely due to technical limitations of the equipment used in that study. (Relevant technical limitations include poor sensitivity of the recording apparatus to low-level TEOAEs, and inadequate rejection or reduction of extraneous noise.) Other studies that showed prevalence figures substantially less than 100% may have been similarly limited. Furthermore, some recent studies (e.g., Prieve *et al.*, 1993) have

involved testing under "typical clinical conditions," rather than ideal laboratory situations, and lower apparent prevalence figures in those studies might be expected due to contamination of weak emissions by extraneous noise.

³Small sample size results in greater uncertainty attached to a conclusion that all normal ears demonstrate TEOAEs, than to a conclusion that some normal ears lack TEOAEs. For example, if the true population prevalence is 97%, the probability of finding emissions in *all* (100%) of a random sample of 50 ears is 0.22 (0.97 raised to the power of 50). Hence, about one in five studies of that size would be expected to show 100% prevalence, purely on statistical grounds. For the probability of obtaining a sample prevalence of 100% to be < 0.05 , when the true population prevalence is (say) 99%, it is necessary to test a random sample of at least 300 ears.

⁴One approach to comparing the performance of different systems of recording and rating is by examination of what is analogous to the sensitivity/specificity trade-off for screening tests. Thus in evaluating the prevalence figures in normals yielded by different systems, one could also compare percentage prevalence of TEOAEs found in *hearing-impaired subjects*. [Prievé *et al.* (1993) address this point in some detail, in a study aimed at optimally separating normal and hearing-impaired subjects under clinical conditions.] To illustrate, Reshef *et al.* (1993) found a 100% prevalence of TEOAEs in their normal ears at click levels of 70 dB SPL and also reported as many as 80% of ears with noise-induced hearing loss registering emissions at this click level. Emissions were identified by subjective rating in their study: More conservative rating could possibly have yielded lower prevalence figures in both subject groups. In contrast, Stevens (1988) found prevalences of 97% in his normals and only 7% in a group of hearing-impaired ears. Although hearing losses were probably more severe in this group than in the hearing-impaired group of Reshef *et al.* (1993), the large difference in prevalence of emissions in impaired ears reported in these two studies may also be a reflection of differences in the subjective rating criteria. (Neither author presents examples of poor emission waveforms from normal ears, which may help address this issue.)

⁵Probst *et al.* (1991) also suggest that in some ears TEOAEs evoked by clicks may be difficult to detect, whereas the use of tone-burst stimuli may result in clearer emissions.

⁶Although a measured TEOAE response may be regarded as containing linear and nonlinear components (Kemp *et al.*, 1986) a convention among many authors (e.g., Grandori *et al.*, 1993; Salomon *et al.*, 1993a) is to denote the responses actually measured (at various stimulus levels) as "linear" responses. The "nonlinear" response is then derived from the "linear" responses at two different stimulus levels.

⁷In fact, a total of 49 of the presumed normal ears that passed on TEOAE classification via r_{nl} also underwent audiometry, and only 3 (6%) of these did not meet the study's audiometric criterion. (These 3 were, of course, excluded in the figure of 397 for our basic test group.)

⁸One qualification on the interpretation of this CI arises because nearly all of our sample ears came from subjects from whom both ears were tested. Many properties of TEOAEs are significantly similar between left and right ears of the same subject (Probst *et al.*, 1986; Bonfils *et al.*, 1988a; Johnsen *et al.*, 1988). However, measures from one ear are not completely dependent on those from the other, and it is difficult to specify the degree of covariation between the two ears. Most studies (including the three cited above and Bonfils *et al.*, 1988b) have pooled the data from the two ears of their subjects. In the case of our sample, only three ears (from three different subjects) failed on r_{nl} classification. Hence, we can make no inferences from these data, concerning associations between TEOAE presence in left ears and in right ears of the same normal subjects. Nevertheless, we may re-estimate our CI to examine the extreme possibility of complete dependence between the two ears by considering results from only left or only right ears. As all three ears which failed on r_{nl} happened to be left, different 95% confidence limits are obtained, namely 96.2% for left and 97.5% for right ears.

⁹Tests of physiological vulnerability of a recorded signal, for example, by exposure of subjects to noise or low doses of aspirin, were beyond the scope of this study. Cochlear origin could also be indicated if the signal could be suppressed by additional acoustic stimulation. However, although we did not attempt such procedures in this study, our experience from other ongoing work is that ipsilateral acoustic stimulation effects little suppression of TEOAEs that are of low level to begin with.

¹⁰Indeed, the original hypothesis of Kemp (1978) was that "impedance discontinuities" arose due to the response of the basilar membrane. Strube (1989, p. 44) and Shera and Zweig (1993, footnote 2) also speculate on TEOAE generation without mechanical irregularities.

¹¹A localized TEOAE generation process (with place-frequency mapping) is implicit under alternative (a). In the case of alternative (b), the generation process for a given TEOAE frequency component may be either local or distributed.

¹²Avan *et al.* (1995) allow for two possible explanations for their data: (a) that TEOAEs are sensitive to damage to basal parts of the cochlea (which correspond to higher characteristic frequencies); or (b) that TEOAEs may indeed be sensitive to damage to those parts *only* of the cochlea which are tuned to their frequencies, but that TEOAEs may be more sensitive to such local damage than are the corresponding auditory thresholds. Our discussion here applies equally to either hypothesis.

Avan, P., Bonfils, P., Loth, D., Elbez, M., and Erminy, M. (1995). "Transient-evoked otoacoustic emissions and high-frequency acoustic trauma in the guinea pig," *J. Acoust. Soc. Am.* **97**, 3012–3020.

Avan, P., Bonfils, P., Loth, D., Narcy, P., and Trotoux, J. (1991). "Quantitative assessment of human cochlear function by evoked otoacoustic emissions," *Hearing Res.* **52**, 99–112.

Avan, P., Bonfils, P., Loth, D., and Wit, H. (1993). "Temporal patterns of transient-evoked otoacoustic emissions in normal and impaired cochleae," *Hearing Res.* **90**, 109–120.

Bonfils, P., Bertrand, Y., and Uziel, A. (1988a). "Evoked otoacoustic emissions: Normative data and presbycusis," *Audiology* **27**, 27–35.

Bonfils, P., Piron, J.-P., Uziel, A., and Pujol, R. (1988b). "A correlative study of evoked otoacoustic emission properties and audiometric thresholds," *Arch. Otorhinolaryngol.* **245**, 53–56.

Bonfils, P., Uziel, A., and Pujol, R. (1988c). "Screening for auditory dysfunction in infants by evoked otoacoustic emissions," *Arch. Otolaryngol. Head Neck Surg.* **114**, 887–890.

Cope, Y., and Lutman, M. E. (1988). "Oto-acoustic emissions," in *Paediatric Audiology 0–5 years*, edited by B. McCormick (Taylor and Francis, London), pp. 221–245.

Dirckx, J. J. J., Daemers, K., Somers, T., Offeciers, F. E., and Govaerts, P. J. (1996). "Numerical assessment of TOAE screening results—currently used criteria and their effect on TOAE prevalence figures," *Arch. Otolaryngol.* **116**, 672–679.

Dolhen, P., Hennaux, C., Chantry, P., and Hennebert, D. (1991). "The occurrence of evoked oto-acoustic emissions in a normal adult population and neonates," *Scand. Audiol.* **20**, 203–204.

Furst, M., and Lapid, M. (1988). "A cochlear model for acoustic emissions," *J. Acoust. Soc. Am.* **84**, 222–229.

Grandori, F., Ravazzani, P., Tognola, G., and Hatzopoulos, S. (1993). "Derived nonlinear" versus "linear" click-evoked otoacoustic emissions," in *Advances in Otoacoustic Emissions: Vol. I—Fundamentals and Clinical Applications*, edited by F. Grandori (Commission of the EC, Brussels), pp. 48–64.

Harris, F. P., and Probst, R. (1991). "Reporting click-evoked and distortion-product otoacoustic emission results with respect to the pure-tone audiogram," *Ear Hear.* **12**, 399–405.

Hauser, R., Probst, R., and Lohle, E. (1991). "Click- and tone-burst-evoked otoacoustic emissions in normally hearing ears and in ears with high-frequency sensorineural hearing loss," *Eur. Arch. Oto-Rhino-Laryngol.* **248**, 345–352.

Johnsen, N. J., Parbo, J., and Elberling, C. (1993). "Evoked acoustic emissions from the human ear. VI. Findings in cochlear hearing impairment," *Scand. Audiol.* **22**, 87–95.

Johnsen, N. J., Bagi, P., Parbo, J., and Elberling, C. (1988). "Evoked acoustic emissions from the human ear. IV. Final results in 100 neonates," *Scand. Audiol.* **17**, 27–34.

Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.* **64**, 1386–1391.

Kemp, D. T. (1986). "Otoacoustic emissions, travelling waves and cochlear mechanisms," *Hearing Res.* **22**, 95–104.

Kemp, D. T., Ryan, S., and Bray, P. (1990). "A guide to the effective use of otoacoustic emissions," *Ear Hear.* **11**, 93–105.

Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (1986). "Acoustic emission cochleography—Practical aspects," *Scand. Audiol.* **25**, 71–95.

Kok, M. R., van Zanten, G. A., Brocaar, M. P., and Wallenberg, H. C. A. (1993). "Click-evoked oto-acoustic emissions (EOAEs) in 1036 ears of healthy newborns," *Audiology* **32**, 213–224.

Lutman, M. E. (1989). "Evoked otoacoustic emissions in adults: implications for screening," *Audiol. Practice* **6**, 6–8.

Lutman, M. E. (1993). "Reliable identification of click-evoked otoacoustic

- emissions using signal-processing techniques," *Br. J. Audiol.* **27**, 103–108.
- Lutman, M. E., and Saunders, G. H. (1992). "Lack of association between otoacoustic emissions and hearing difficulty in subjects with normal hearing thresholds," *J. Acoust. Soc. Am.* **92**, 1184–1185.
- Lutman, M. E., Jennings, K., Davis, A. C., Houston, H. G., and Meredith, R. (1993). "Coloration of click-evoked otoacoustic emissions by characteristics of the recording apparatus," in *Advances in Otoacoustic Emissions: Vol. I—Fundamentals and Clinical Applications*, edited by F. Grandori (Commission of the EC, Brussels), pp. 36–47.
- Prieve, B. A., Gorga, M. P., Schmidt, A. R., Neely, S. T., Peters, J., Schultes, L., and Jesteadt, W. (1993). "Analysis of transient-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," *J. Acoust. Soc. Am.* **93**, 3308–3319.
- Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (1991). "A review of otoacoustic emissions," *J. Acoust. Soc. Am.* **89**, 2027–2067.
- Probst, R., Coats, A. C., Martin, G. K., and Lonsbury-Martin, B. L. (1986). "Spontaneous click- and toneburst-evoked otoacoustic emissions from normal ears," *Hearing Res.* **21**, 261–275.
- Reshef, I., Attias, J., and Furst, M. (1993). "Characteristics of click-evoked otoacoustic emissions in ears with normal hearing and with noise-induced hearing loss," *Br. J. Audiol.* **27**, 387–395.
- Rutten, W. L. C. (1980). "Evoked acoustic emissions from within normal and abnormal human ears: Comparison with audiometric and electrocochleographic findings," *Hearing Res.* **2**, 263–271.
- Salomon, G., Groth, J., and Anthonisen, B. (1993b). "Preliminary results and considerations in hearing screening of newborns based on otoacoustic emissions," *Br. J. Audiol.* **27**, 139–141.
- Salomon, G., Anthonisen, B., Groth, J., Pape-Thomsen, P., Berget, A., Olesen, H., Parbo, J., Kofod-Hansen, B., Liang, C. F., and Bloch, M. (1993a). "Pilot study preceding a systematic hearing screening of newborns based on evoked otoacoustic emissions," in *Advances in Otoacoustic Emissions: Vol. I—Fundamentals and Clinical Applications*, edited by F. Grandori (Commission of the EC, Brussels), pp. 131–140.
- Shera, C. A., and Zweig, G. (1993). "Noninvasive measurement of the cochlear travelling-wave ratio," *J. Acoust. Soc. Am.* **93**, 3333–3352.
- Stevens, J. C. (1988). "Click-evoked oto-acoustic emissions in normal and hearing-impaired adults," *Br. J. Audiol.* **22**, 45–49.
- Stevens, J. C., Webb, H. D., Smith, M. F., Buffin, J. T., and Ruddy H. (1987). "A comparison of oto-acoustic emissions and brain stem electric response audiometry in the normal newborn and babies admitted to a special care baby unit," *Clin. Phys. Physiol. Meas.* **8**, 95–104.
- Strube, H. W. (1989). "Evoked otoacoustic emissions as cochlear Bragg reflections," *Hearing Res.* **38**, 35–46.
- Sutton, G. J., and Wilson, J. P. (1983). "Modelling cochlear echoes: The influence of irregularities in frequency mapping on summed cochlear activity," in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever (Delft University Press, Delft), pp. 83–90.
- Tognola, G., Ravazzani, P., and Grandori, F. (1995). "An optimal filtering technique to reduce the influence of low-frequency noise on click-evoked otoacoustic emissions," *Br. J. Audiol.* **29**, 153–160.
- van Dijk, P., and Wit, H. P. (1987). "The occurrence of click-evoked otoacoustic emissions (Kemp echoes) in normal-hearing ears," *Scand. Audiol.* **16**, 62–64.
- Vedantam, R., and Musiek, F. E. (1991). "Click evoked otoacoustic emissions in adult subjects: standard indices and test-retest reliability," *Am. J. Otol.* **12**, 435–442.
- White, K. R., Vohr, B. R., Maxon, A. B., Behrens, T. R., and Mauk, G. W. (1994). "Screening all newborns for hearing loss using transient evoked otoacoustic emissions," *Am. J. Paed. Otolaryngol.* **29**, 203–217.
- Wilson, J. P. (1980). "Model for cochlear echoes and tinnitus based on an observed electrical correlate," *Hearing Res.* **2**, 527–532.
- Wit, H. P., van Dijk, P., and Avan, P. (1994). "Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions," *Hearing Res.* **73**, 141–147.

Changes in evoked otoacoustic emissions in the guinea pig after pure-tone acoustic overstimulation

Hiromi Ueda, Hayato Tsuge, and Taku Hattori

Department of Otolaryngology, Nagoya University, School of Medicine, 65 Tsurumai-Cho, Showa-Ku, Nagoya, 466 Japan

(Received 6 March 1996; revised 10 February 1997; accepted 25 February 1997)

To test if click-evoked otoacoustic emissions (CEOAEs) have frequency specificity, continuous changes in CEOAEs (especially frequency components of the CEOAE power spectrum) after pure-tone exposure in guinea pigs were examined. Pure-tone stimuli (0.5 kHz, 120 dB SPL; 2 kHz, 115 dB SPL; 4 kHz, 110 dB SPL) were given in a closed system for 3 min. After exposure, the frequency components in the CEOAE power spectrum decreased maximally at one-half octave or more above the overstimulation frequency. They partially recovered 2 h after exposure. The time course of compound action potential (CAP) thresholds after exposure was similar to that of the frequency components of the CEOAE power spectrum. It was concluded that some local damage caused by outer hair-cell dysfunction in the guinea pig cochlea can be detected by measuring shifts in frequency components in the CEOAE power spectrum. © 1997 Acoustical Society of America. [S0001-4966(97)06006-2]

PACS numbers: 43.64.Jb, 43.64.Nf, 43.64.Wn [RDF]

INTRODUCTION

Otoacoustic emissions (OAEs) are reflections of cochlear micromechanics from the inner ear to the middle ear and external ear canal. OAEs are thought to be linked with the function of outer hair cells, and to reflect active processes in the cochlea. In addition, if acoustic emissions indeed reflect the state of a specific region along the cochlea, they will be useful as a research or clinical tool for noninvasive assessment of cochlear function. One kind of acoustic emission, distortion product otoacoustic emission (DPOAE) is thought to be frequency specific in both clinical research (Avan and Bonfils, 1993) and animal studies (Schmiedt, 1986; Martin *et al.*, 1987), when $2f_1-f_2$ are measured. Recently, Kemp *et al.* (1990) and Norton (1993) have suggested that a click evoked otoacoustic emission (CEOAE) component of a particular frequency arises from the part of the cochlea tuned to that frequency in humans. On the other hand, several researchers (Wilson, 1980; Avan *et al.*, 1991a) have assumed that CEOAEs result from the summed contribution of mechanisms distributed along the basilar membrane and have had some doubts about the frequency specificity of CEOAEs.

Transiently evoked otoacoustic emissions (TEOAEs) have been difficult to detect in guinea pigs, as reported by Zwicker and Manley (1981). Use of the otodynamic analyzer ILO 88, devised by Bray and Kemp (1987), has made measurement of TEOAEs (include CEOAEs) in guinea pigs both easy and consistent (Ueda *et al.*, 1992; Ross *et al.*, 1994). In this study, we measured changes in CEOAEs in guinea pigs after pure-tone acoustic overstimulation which can cause a temporary local injury to the cochlea. By comparing frequency components in the CEOAE frequency power spectrum before and after exposure, we tested whether CEOAEs are useful as a tool for detecting local damage to the guinea pig cochlea.

I. MATERIALS AND METHODS

Healthy, white guinea pigs about 200 g in size were anesthetized with pentobarbital, placed in soundproof chamber, and restrained with a head holder. Tracheotomy and catheterization of the jugular vein were performed under spontaneous breathing. Body temperature was maintained at 37 °C throughout the experiment. After the cartilaginous external ear canal was cut away, a customized metal tube was inserted into the ear canal. For measurement of compound action potentials (CAPs) and pure-tone exposure, a commercial ear phone (Sony MDR-E 565, Tokyo) was attached to the metal tube. A silver electrode was placed on the auditory bulla near the styloid process as a negative input. A positive input was placed at the test ear mastoid. A tone burst which was 1 ms in rise-fall decay time and 3 ms in duration time was delivered through the earphone at a rate of 10/s, and transient responses were averaged 100 times. The visual detection threshold level was measured via 5-dB decrements. Measurement of CEOAEs utilized an otodynamic analyzer ILO 88. An E-type neonate OAE probe for the ILO 88 was attached to the metal tube instead of the ear phone. Transient responses to nonlinear clicks at about 70 dB pe SPL were averaged 260 times and analyzed during the first 20 ms after stimulus onset. Generally, acoustic ringing in individual ear canals decreases in amplitude as time elapses and disappears after 1 ms at 70 dB pe SPL. On the other hand, since CEOAEs increase in amplitude from 1 ms after stimulus onset, we can easily detect CEOAEs visually. After CEOAEs were distinguished from acoustic ringing in individual ear canals, the time window was then set from 1 to 10 ms after stimulus onset. Then, the original waveform was reconstructed. From total echo energy (total echo power, TEP) in the cross-power spectrum of the two independent echo responses, a more precise analysis was conducted after conversion into frequency bands of 1000 Hz using a frequency filter in ILO 88: Five frequency bands (filtered echo

power, FEP) between 0.5 and 5.5 kHz were analyzed. Bands were composed as follows: FEP1=0.5–1.5 kHz; FEP2=1.5–2.5 kHz; FEP3=2.5–3.5 kHz; FEP4=3.5–4.5 kHz; and FEP5=4.5–5.5 kHz.

CEOAEs were measured every 10 min for 30 min, and CAPs were measured twice. When responses were considered stable, the animals were exposed to pure tones in a closed system for 3 min. We divided the animals into four groups, each exposed to a different frequency pure tone. As a control, we first measured CEOAEs without any pure-tone exposure. Pure tones were 0.5 kHz at 120 dB SPL, 2 kHz at 115 dB SPL, and 4 kHz at 110 dB SPL. Calibration of the sound-pressure level was performed through the hole made in the external ear canal using the B & K probe microphone (type 4182). CEOAEs were measured every 10 min until 2 h after the exposure. CAPs were measured every 30 min, simultaneous with the measurement of CEOAEs. In the groups exposed to 0.5 kHz, 2 kHz, and 4 kHz, we measured CAPs at 1 kHz, 3 kHz, and 6 kHz, respectively. Data were calculated and reported as the difference between before and after exposure. Repeated measure ANOVA was used to compare the extent of changes among individual frequency band data and those between pre- and post-exposure. After ANOVA, means comparisons test were performed using contrast method. A level of $P < 0.05$ was accepted as statistically significant.

The investigations have been performed in accordance with the principles of the Declaration of Helsinki.

II. RESULTS

Throughout the experiment, background noise in CEOAEs measurements never exceeded 5 dB SPL for TEP, –2 dB SPL for FEP1, –5 dB SPL for FEP2 and FEP3, and –10 dB SPL for FEP4 and FEP5 measurements, respectively.

A. Control experiment

Figure 1 shows the time course of TEP and each FEP without any exposure. TEP and each FEP were stable throughout the experiment, and there were no significant differences among them.

B. 0.5-kHz exposure

Figure 2 changes in 1-kHz CAP thresholds, TEP, and each FEP before and after 0.5-kHz pure-tone exposure. The degree of echo power change differed greatly for each frequency component immediately after the exposure. FEP1 showed the greatest reduction. On the other hand, recovery of each FEP from 30 min to 120 min after the exposure was similar among the three FEPs. Changes in 1-kHz CAP thresholds after exposure were the same tendency as those of FEP1. When all measurements of CAP and FEP1 at all measurement time points were pooled, the increase in CAP threshold at 1 kHz was plotted as a function of FEP1, and it appeared to follow a linear-regression law:

$$\Delta\text{CAP}(\text{dB}) = 0.921 \times \Delta\text{FEP1} - 1.826, \quad r = 0.605.$$

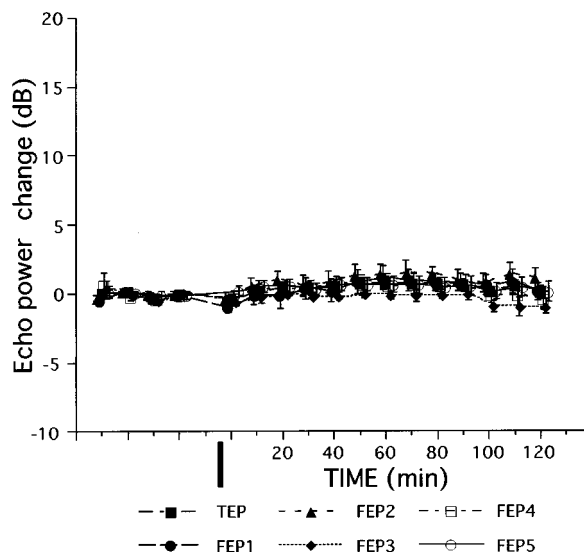


FIG. 1. Time course of shifts in total echo power (TEP) and individual filtered echo powers (FEPs) without any pure-tone exposure. We normalized the values and calculated the mean values of the control period to be 0 dB. Vertical bar indicates one standard error. $N=6$.

This regression was significant ($P < 0.01$). Figure 3 shows shifts in FEP levels immediately and 2 h after exposure. The reductions in FEP1 and FEP2 were statistically significant immediately after exposure (Table I). Any other FEP did not change significantly after exposure.

C. 2-kHz exposure

Figure 4 shows the time course of the FEPs and 3-kHz CAP thresholds before and after 2-kHz pure-tone exposure. FEP2 and FEP3 showed the greatest reduction after exposure, followed by slow recovery. However, they did not reach pre-exposure values even 2 h after exposure. FEP5 showed only minimal change. The time course of 3-kHz CAP thresholds after exposure was the same tendency as that of FEP2 and FEP3. These also failed to reach pre-exposure levels even 2 h after the exposure. When plotted as a function of FEP3, the increase in CAP threshold at 3 kHz appeared to follow a linear-regression law:

$$\Delta\text{CAP}(\text{dB}) = 1.904 \times \Delta\text{FEP3} + 1.829, \quad r = 0.841.$$

This regression was highly significant ($P < 0.0001$). Figure 5 shows shifts in FEP levels immediately and 2 h after exposure. FEP2 and FEP3 showed maximum reduction at both time points. Although the reductions in all FEPs were statistically significant immediately after exposure, those in FEP1 and FEP5 became insignificant 2 h after exposure (Table II).

D. 4-kHz exposure

Figure 6 shows the time course of the FEPs and 6-kHz CAP thresholds before and after 4-kHz pure-tone exposure. FEP5 showed the greatest reduction after exposure, and the time course of that was the same tendency as that of 6-kHz CAP thresholds. Neither indicator recovered pre-exposure

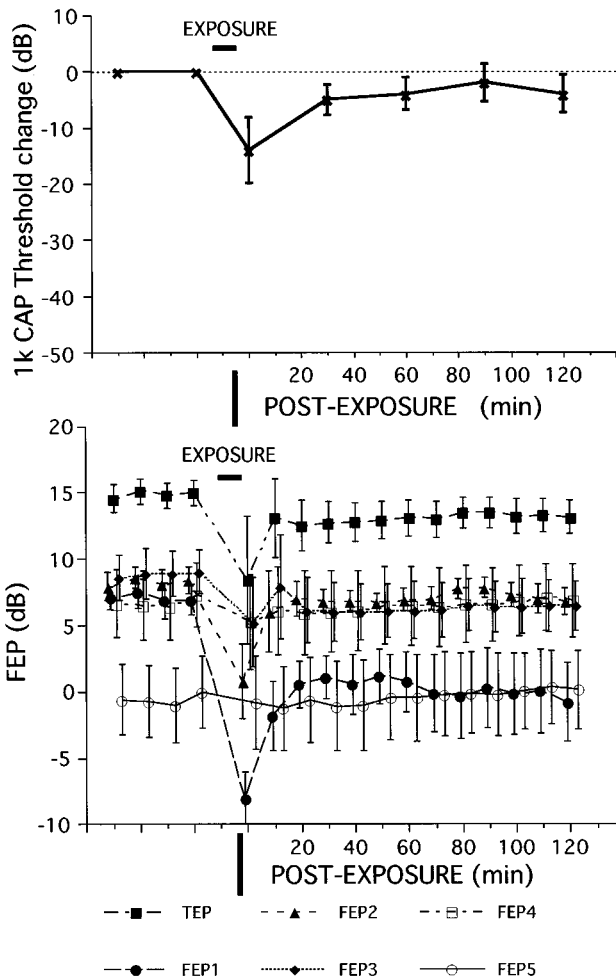


FIG. 2. Changes in 1-kHz compound action potential (CAP) thresholds, total echo power (TEP), and individual filtered echo powers (FEPs) before and after 0.5-kHz pure-tone exposure (120 dB SPL, 3 min). Upper figures shows changes in 1-kHz CAP threshold. Lower figure shows changes in TEP and individual FEPs. Vertical bar indicates one standard error. We calculated that no echo powers were smaller than -5 dB (TEP) or -10 dB (FEP). $N=5$.

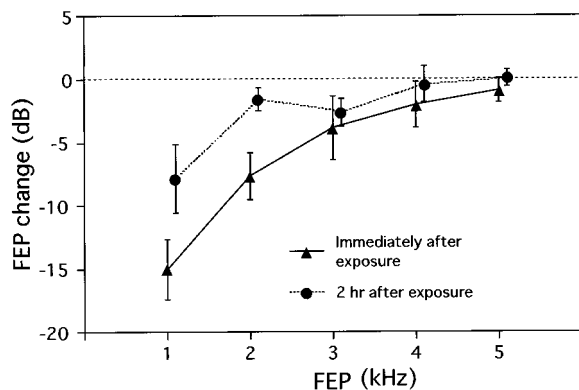


FIG. 3. Shifts in filtered echo power (FEP) levels immediately and 2 h after 0.5-kHz pure-tone exposure (120 dB SPL, 3 min). We normalized the values and calculated the mean values of the control period to be 0 dB. Vertical bar indicates one standard error. $N=5$.

TABLE I. Significance of the difference of values in individual filtered echo power (FEP) levels before vs after 0.5-kHz pure-tone exposure (120 dB SPL, 30 min). $N=5$.

| | FEP1 | FEP2 | FEP3 | FEP4 | FEP5 |
|---------|-----------|----------|------|------|------|
| 0 min | $P<0.001$ | $P<0.05$ | NS | NS | NS |
| 120 min | $P<0.05$ | NS | NS | NS | NS |

values, even 2 h after the exposure. When plotted as function of FEP5, the increase in CAP threshold at 6 kHz appeared to follow a linear-regression law:

$$\Delta\text{CAP}(\text{dB})=3.079\times\Delta\text{FEP5}-0.43, \quad r=0.876.$$

This regression was highly significant ($P<0.0001$). Figure 7 shows that the reduction in FEP5 was greatest among the five FEPs both immediately and 2 h after the exposure. The reductions in FEP4 and FEP5 were statistically significant immediately after exposure (Table III). No other FEP changed significantly.

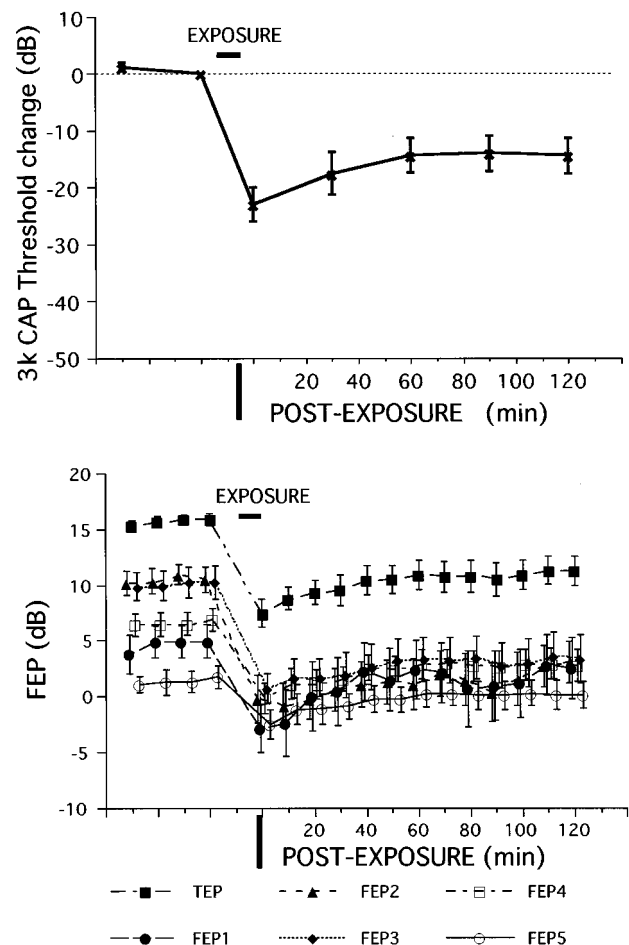


FIG. 4. Changes in 3-kHz compound action potential (CAP) thresholds, total echo power (TEP) and individual filtered echo powers (FEPs) before and after 2-kHz pure-tone exposure (115 dB SPL, 3 min). Upper figure shows changes in the 3-kHz CAP threshold. Lower figure shows changes in TEP and individual FEPs. Vertical bar indicates one standard error. We calculated that no echo powers were smaller than -5 dB (TEP) or -10 dB (FEP). $N=8$.

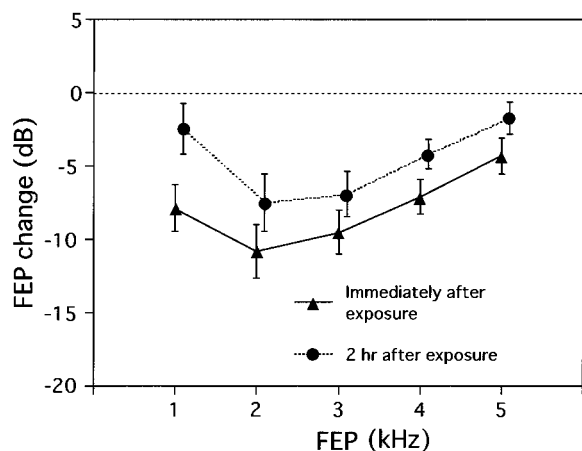


FIG. 5. Shifts in FEP levels immediately and 2 h after 2-kHz pure-tone exposure (115 dB SPL, 3 min). We normalized the values and calculated the mean values of the control period to be 0 dB. Vertical bar indicates one standard error. $N=8$.

III. DISCUSSION

OAEs are now popularly used as a research or clinical tool for noninvasive assessment of the active process, as proposed by Davis (1983). One of the most interesting issues is whether OAEs can indeed reflect the state of a specific region along the cochlea. Among OAEs, DPOAEs and CEOAEs are the leading candidates for detecting local damage along the cochlea.

For DPOAEs, some authors (Avan and Bonfils, 1993; Lonsbury-Martin and Martin 1990) have concluded that DPOAEs at $2f_1$ - f_2 have frequency specificity because of good correlation between auditory thresholds and geometric mean values of the two primary frequencies. On the other hand, there is some question as to whether TEOAEs can reflect activity of a specific region along the cochlea. Some authors (Kemp *et al.*, 1990; Norton, 1993) have suggested that CEOAEs can be used to identify frequency ranges for normal hearing in ears with pathology by measuring frequency components in the CEOAE response spectrum. Collet *et al.* (1991) reported significant correlation between spectrum analysis of CEOAEs and hearing losses in pure-tone audiograms. Gorga *et al.* (1993) examined CEOAEs analyzed into octave or one-third octave bands for frequencies ranging from 500 to 4000 Hz to distinguish normal hearing from hearing impairment. They concluded that CEOAEs were able to separate normal from impaired ears at 1 kHz, 2 kHz, and 4 kHz. But they did not mention correlation between audiometric threshold and CEOAEs. To create local damage to the active process in the cochlea, acoustic overstimulation was used by some investigators (Rossi *et al.*,

TABLE II. Significance of the difference of values in individual filtered echo power (FEP) levels before versus after 2-kHz pure-tone exposure (115 dB SPL, 3 min). $N=8$.

| | FEP1 | FEP2 | FEP3 | FEP4 | FEP5 |
|---------|----------|-----------|-----------|-----------|----------|
| 0 min | $P<0.05$ | $P<0.001$ | $P<0.001$ | $P<0.001$ | $P<0.05$ |
| 120 min | NS | $P<0.01$ | $P<0.01$ | $P<0.05$ | NS |

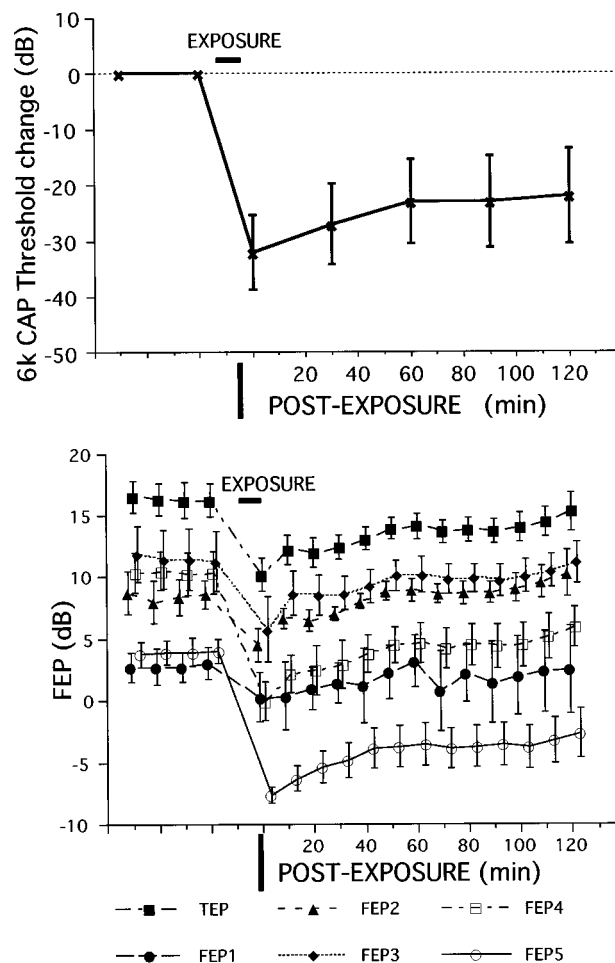


FIG. 6. Changes of 6-kHz compound action potential (CAP) thresholds, total echo power (TEP), and individual filtered echo powers (FEPs) before and after 4-kHz pure-tone exposure (110 dB SPL, 3 min). Upper figure shows changes in the 6-kHz CAP threshold. Lower figure shows changes in TEP and individual FEPs. Vertical bar indicates one standard error. We calculated that no echo powers were smaller than -5 dB (TEP) or -10 dB (FEP). $N=5$.

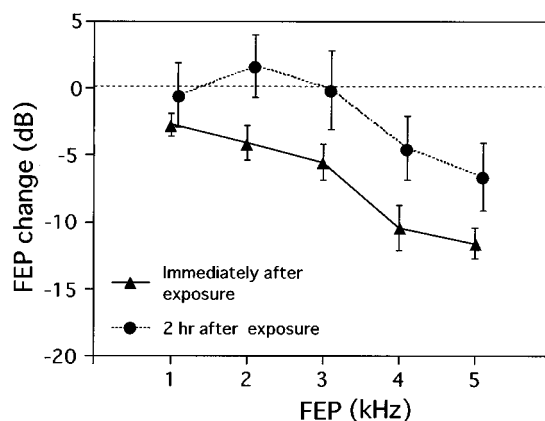


FIG. 7. Shifts in filtered echo power (FEP) levels immediately and 2 h after 4-kHz pure-tone exposure (110 dB SPL, 3 min). We normalized the values and calculated the mean values of the control period to be 0 dB. Vertical bar indicates one standard error. $N=5$.

TABLE III. Significance of the difference of values in individual filtered echo power (FEP) levels before vs after 4-kHz pure-tone exposure (110 dB SPL, 3 min). $N=5$.

| | FEP1 | FEP2 | FEP3 | FEP4 | FEP5 |
|---------|------|------|------|-----------|------------|
| 0 min | NS | NS | NS | $P<0.001$ | $P=0.0001$ |
| 120 min | NS | NS | NS | NS | $P<0.01$ |

1991; Wilson and Evans, 1983; Avan *et al.*, 1991b). In this experiment, we also exposed guinea pigs to pure-tone overstimulation. To assess local pathology, we divided the CEOAE response spectrum into five frequency components from 1 kHz to 5 kHz. Collet *et al.* (1991) also divided the CEOAE response spectrum, but into more components. We thought fragmentation of the CEOAE frequency components could generate artifact because frequency filters are not rectangular and therefore adopted frequency bands of 1000 Hz.

It is well documented that the maximum damage site after pure-tone exposure is at one-half octave above the overstimulation frequency (Mitchell *et al.*, 1977; Davis, 1983). For this reason, we measured continuous changes in CAP at about one-half octave above the overstimulation frequency. Our results showed that the frequency components in the CEOAE power spectrum were maximally reduced at one-half octave or more above the overstimulation frequency, and that the time course of the FEPs was similar to those of the CAPs. In Fig. 5, the maximum FEP changes seemed to occur at 2 kHz, but the FEP changes at 3 kHz were also as prominent as those at 2 kHz. Although we did not measure the CAP changes at 2 kHz, the CAP changes at 2 kHz might be almost equal to those at 3 kHz. In addition, FEP changes at 4 kHz were larger than those at 1 kHz, indicating that the damage after pure-tone exposure was more severe above the overstimulation frequency. There was a strong correlation between CAP at 3 kHz and FEP3 and between CAP at 6 kHz and FEP5. On the other hand, the correlation between CAP at 1 kHz and FEP1 was weak although significant. Two main reasons can be considered. First, background noise is most likely to occur in the low frequencies. Prieve *et al.* (1993) also mentioned this idea and suggested one of the reasons why audiometric threshold at 500 Hz were not related to CEOAEs. Second, our method for measurement of CAP used a random-phased trigger to cancel the cochlear microphonics. Eggermont (1976) indicated that CAP at 500 Hz and 1 kHz appeared during a fixed phase of the sine wave in the cochlear microphonics. The use of the unfixed phased trigger may cause relatively high CAP thresholds at 1 kHz in the pre-exposure state and induce smaller changes in the 1-kHz CAP threshold after exposure in spite of the high intensity stimulus. In regression laws, the slope of FEP changes versus CAP threshold shifts depended on exposure frequency. One reason is the use of the unfixed phased trigger as mentioned above. Another reason may be that inner hair cells and/or their synapses with auditory nerve dendrites are damaged after exposure. Rossi *et al.* (1991) examined temporal changes in behavioral thresholds of 1-kHz tone bursts and thresholds of 1-kHz tone burst evoked OAEs after 0.75-kHz pure-tone exposure in man, revealing a similarity between characteristics and modalities of recovery time of

both thresholds. Wilson and Evans (1983) reported that 2.33-kHz tone burst evoked OAEs in cats decreased reversibly after 1.8-kHz pure-tone exposure but were little affected after 5-kHz pure-tone exposure. These results agree with our results. However, Avan *et al.* (1991b) revealed that tone burst evoked OAEs decreased in proportion to temporary threshold shifts in CAPs at one-half octave above the exposure frequency, but decreased even without CAP decreases at a lower frequency than the exposure tone in guinea pigs. They suggested that TEOAEs are sensitive to the global cochlear state on the high-frequency side rather than to the local hearing threshold shifts. They also suggested the same finding was true in man after measuring latency shortening of impaired ears (Avan *et al.*, 1993). These findings differ from our results and those reported by Wilson and Evans (1983). Recently Avan *et al.* (1995) investigated relationships between CAP thresholds and CEOAEs before and after pure-tone exposure at relatively high frequencies, and reported that CEOAE components changed slightly at lower frequencies at which CAP thresholds did not change. They indicated two possibilities: acoustic trauma induced minute cochlear damage at places tuned to low frequencies and CEOAE was more sensitive to it than CAP threshold; significant contributions to lower-frequency CEOAE came from the most damaged cochlear places tuned to much higher frequencies. In Fig. 7 FEP1, FEP2, and FEP3 seemed to reduce their amplitude immediately after exposure, although there were no significant statistically. In this experiment, we can suppose that any CAP thresholds at those frequencies would not have changed, although we did not measure those CAP thresholds. This means that CEOAE components at lower frequencies reduced slightly without CAP threshold changes at those frequencies. However, FEP1, FEP2, and FEP3 recovered to pre-exposure level 2 h after exposure. In addition, echo powers exceeded the pre-exposure level especially at 2 kHz in some animals. From these results, we speculate that there were some minute cochlear damage at places tuned to low frequencies immediately after 4-kHz exposure.

In conclusion, pure-tone exposure damaged the active process in the guinea pig cochlea, and CEOAEs are frequency specific in guinea pigs.

- Avan, P., and Bonfils, P. (1993). "Frequency specificity of human distortion product otoacoustic emissions," *Audiology* **32**, 12–26.
- Avan, P., Loth, D., Menguy, C., and Teyssou, M. (1991b). "Frequency dependence of changes in guinea-pig cochlear emissions after acoustic overstimulation," *J. Acoust. Soc. Am.* **4**, 91–94.
- Avan, P., Bonfils, P., Loth, D., and Wit, H. P. (1993). "Temporal patterns of transient-evoked otoacoustic emissions in normal and impaired cochleae," *Hearing Res.* **70**, 109–120.
- Avan, P., Bonfils, P., Loth, D., Elbez, M., and Erminy, M. (1995). "Transient-evoked otoacoustic emissions and high-frequency acoustic trauma in the guinea pig," *J. Acoust. Soc. Am.* **97**, 3012–3020.
- Avan, P., Bonfils, P., Loth, D., Narcy, P., and Trotoux, J. (1991a). "Quantitative assessment of human cochlear function by evoked otoacoustic emissions," *Hearing Res.* **52**, 99–112.
- Bray, P., and Kemp, D. T. (1987). "An advanced cochlear echo technique suitable for infant screening," *Br. J. Audiol.* **21**, 191–204.
- Collet, L., Veuillet, E., Chanal, J. M., and Morgan, A. (1991). "Evoked otoacoustic emissions: Correlates between spectrum analysis and audiogram," *Audiology* **30**, 164–172.
- Davis, H. (1983). "An active process in cochlear mechanics," *Hearing Res.* **9**, 79–90.

- Eggermont, J. J. (1976). "Basic principles for electrocochleography," *Acta Oto-Laryngol. Suppl.* **316**, 7–16.
- Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (1993). "A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects," *J. Acoust. Soc. Am.* **94**, 2639–2648.
- Kemp, D. T., Ryan, S., and Bray, P. (1990). "A guide to the effective use of otoacoustic emissions," *Ear. Hear.* **11**, 93–105.
- Lonsbury-Martin, B. L., and Martin, G. K. (1990). "The clinical utility of distortion-product otoacoustic emissions," *Ear Hear.* **11**, 144–154.
- Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (1987). "Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures," *Hearing Res.* **28**, 191–208.
- Mitchell, C., Brummett, R., and Vernon, J. (1977). "Frequency effects of temporary N1 depression following acoustic overload," *Acta Oto-Laryngol.* **103**, 117–123.
- Norton, S. J. (1993). "Application of transient evoked otoacoustic emissions to pediatric populations," *Ear. Hear.* **14**, 64–73.
- Prieve, B. A., Gorga, M. P., Schmidt, A., Neely, S. T., Peters, J., Schultes, L., and Jesteadt, W. (1993). "Analysis of transient-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," *J. Acoust. Soc. Am.* **93**, 3308–3319.
- Ross, H. U., Rogowski, M., Reiss, G., and Gloddek, B. (1994). "Detection of cochlear dysfunction by the measurement of transiently evoked otoacoustic emissions in guinea pigs with autoimmune-induced labyrinthitis," *Eur. Arch. Otolaryngol.* **251**, 80–83.
- Rossi, G., Solero, P., Rolando, M., and Olina, M. (1991). "Recovery time of the temporary threshold shift for delayed evoked otoacoustic emissions and tone bursts," *J. Otorhinolaryngol. Relat. Spec.* **53**, 15–18.
- Schmiedt, R. A. (1986). "Acoustic distortion in the ear canal. I. Cubic difference tones: Effects of acute noise injury," *J. Acoust. Soc. Am.* **79**, 1481–1490.
- Ueda, H., Hattori, T., Sawaki, M., Niwa, H., and Yanagita, N. (1992). "The effect of furosemide on evoked otoacoustic emissions in guinea pigs," *Hearing Res.* **62**, 199–205.
- Wilson, J. P. (1980). "Model for cochlear echoes and tinnitus based on an observed electrical correlate," *Hearing Res.* **2**, 527–532.
- Wilson, J. P., and Evans, E. G. (1983). "Effects of furosemide, flaxedil, noise and tone over-stimulation on the evoked otoacoustic emissions in cat," *Proc. Int. Union Physiol. Sci.* **15**, 100.
- Zwicker, E., and Manley, G. (1981). "Acoustical responses and suppression-period patterns in guinea pigs," *Hearing Res.* **4**, 43–52.

The mechanical waveform of the basilar membrane. I. Frequency modulations (“glides”) in impulse responses and cross-correlation functions^{a)}

Egbert de Boer^{b)}

Room D2-226, Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Alfred L. Nuttall^{c)}

Oregon Hearing Research Center, NRC04, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201-3098 and Kresge Hearing Research Institute, University of Michigan, 1301 E. Ann Street, Ann Arbor, Michigan 48109-0506

(Received 14 March 1996; revised 9 December 1996; accepted 7 February 1997)

The purpose of this investigation is to present evidence from experimental as well as model results on temporal variations of the frequency of oscillation in the basilar membrane's impulse response. Stimuli were either clicks leading to a *direct* estimate of the impulse response, or bands of pseudo-random noise (one or two octaves wide) which lead to an *indirect* estimate of the impulse response via a cross-correlation procedure. The noise bands were centered at the best frequency of the BM location under observation. Responses were obtained from the basal turn of the guinea-pig cochlea, from a location with a best frequency (for the weakest stimuli) between 17.0 and 18.5 kHz. Data acquisition was done with a sample frequency of 208 kHz. Input-output cross-correlation functions were found to share with impulse responses the property that the initial oscillations have a noticeably lower frequency than the later ones. During the impulse response the frequency of oscillation increases gradually. This increase occurs and continues to beyond the time that the oscillations reach the largest amplitude. This frequency variation is called a “glide.” Using the “analytic signal” method the frequency of oscillation is found to increase continually throughout the duration of the main lobe of oscillation, even at the lowest tested stimulus intensities (about 20 dB SPL). At high stimulus intensity both the direct and indirect impulse response change their appearance drastically but the glide retains its basic form. In the case of the direct impulse response estimate the glide can be attributed to temporal variation of the degree of nonlinearity. For the indirect impulse response this is not true, because with a constant level noise stimulus there is no regular temporal variation of nonlinearity. In this case the glide should be interpreted as an intrinsic property of the cochlear system. From our and others' data the glide was found to exist over a topographic frequency range of best frequencies of at least from 1.76 to 18 kHz. Two examples of present-day models of the cochlea are discussed of which one is found to demonstrate the glide phenomenon in its response, and the other one does not. © 1997 Acoustical Society of America. [S0001-4966(97)04506-2]

PACS numbers: 43.64.Kc, 43.64.Bt [RDF]

GENERAL INTRODUCTION

As a biomechanical device, the cochlea shows an impressive degree of frequency selectivity. With the somewhat artificial types of stimulus (such as pure tones) that are usually employed in physiological experiments in the ear, the cochlea shows noticeable nonlinear behavior, and the nonlinearity is also found to depend strongly on frequency. In the actual habitat of animals most acoustical stimuli generally are not of the special types as used in physiological experiments and the same is true for the most important acoustical stimulus for man: speech. Typical “natural” sounds and speech signals have been used in physiological experiments

on neural coding (starting with Aertsen and Johannesma, 1980; Aertsen *et al.*, 1980, 1981; Sachs and Young, 1979, 1980, and continuing until the present day) but the precise role played by cochlear distortion under such more general conditions has not precisely been analyzed. The same is true for understanding what processes are occurring, physically and physiologically, under such stimulation conditions. In the case of responses to impulse sounds, our understanding of cochlear functioning seems to be only marginally better than for sinusoidal stimuli, but this understanding is only qualitative, to say the least. Because stationary random (noise-like) stimuli form a bridge between sinusoidal and impulse stimuli on the one hand, and natural signals on the other, these signals may be of great value as tools for finding out the relative importance of nonlinear effects in the waveform of basilar-membrane movement.

In the present report data are presented on the mechani-

^{a)}Part of this material has been presented at the 1995 Midwinter meeting of the ARO (abstract 747).

^{b)}Electronic mail: e.deboer@amc.uva.nl

^{c)}Electronic mail: nuttall@ohsu.edu

cal response of one location on the basilar membrane (BM) in the basal turn of the guinea-pig cochlea to various kinds of stimuli, including clicks and random-noise signals comprising a band of frequency components which is much wider than the passband of the location of measurement. The main goal of these experiments is to provide a common framework to describe the nature of linear and nonlinear aspects of the response to such stimuli. The emphasis in this paper is on temporal phenomena, such as impulse responses, but responses to stationary signals are also interpreted in the form of equivalent impulse responses. Thus, we describe properties of *direct* (from impulsive stimuli) and *indirect* (from noise stimuli) mechanical impulse responses of the cochlea.

In both direct and indirect estimates of the impulse response the initial oscillations are found to have a noticeably lower frequency than the later ones, and the frequency of oscillation gradually increases with time. This increase occurs and continues during and past the time that the oscillations reach the largest amplitude. We have termed this frequency variation a “glide.” We found this phenomenon at all levels of stimulation used. Robles *et al.* (1976) show evidence of it in (high-level) impulse responses measured with the Mössbauer method in the squirrel monkey. A similar phenomenon characterizes results described by Møller (1983). We quantified the glide by way of the “analytic-signal” technique. At high stimulus intensity the impulse response changes its appearance significantly; its oscillations appear to start earlier and to rise faster in amplitude. Since we can quantify the glide only in the period where the oscillations are maximal the glide appears to be shorter than at low stimulus levels. In the case of the direct impulse response estimate the glide can be attributed to temporal variation of the degree of nonlinearity. However, for the indirect impulse response this is not true, because with noise stimulation there is no regular temporal variation of nonlinearity. Therefore, the glide should be interpreted as an intrinsic property of the cochlear system. We collected evidence about the same phenomenon in data from experiments performed by other authors.

The scope of this paper does not permit us to discuss theoretical aspects of the origin of the glide too deeply. However, we briefly discuss two examples of present-day models of the cochlea of which one demonstrates the glide phenomenon, and the other one does not. This finding indicates that there is currently little or no theoretical basis for understanding the glide phenomenon.

I. DIRECT AND INDIRECT IMPULSE RESPONSES

In low-level stimulus situations the relative amount of cochlear nonlinearity is small. In those cases the mechanical transduction of the cochlea can be adequately described by an *impulse response function* (as if the system were linear). This function can be measured *directly*, i.e., by presenting an impulsive acoustical stimulus and recording the mechanical response, or *indirectly*, by presenting a wide-band noise stimulus and computing the cross-correlation function between stimulus and response signals (see Appendix A). Theoretically, when the system is linear or approximately linear, both methods lead to the same result. In the case of

the cochlea, this is true when the peak acoustical stimulus level is below approximately 40 dB SPL. At higher levels the two estimates of the impulse response start to differ because nonlinear effects are more pronounced and express themselves differently in the impulse response than in the cross-correlation function.

With intense impulsive stimuli, the cochlea is partly saturated (overloaded) in the first part of the response, and thus has a lower best frequency (BF); later on, when much of the initial energy has been dissipated, the response gradually becomes that of the nonsaturated cochlea—which is tuned to a higher frequency. If this interpretation is correct, the directly measured impulse response reflects a degree of nonlinearity that varies with time. On the other hand, the indirect impulse response is a good descriptor of cochlear transduction (a kind of “system function”) in cases where the ear is stimulated with stationary wide-band, noise-like, signals, and this is true for all levels of stimulation. In essence, this statement implies that with stimulation by random noise the system is linearized. What this means is formalized in the EQ-NL theorem (de Boer, 1997) which relates the indirect impulse response of a nonlinear cochlear model to the (indirect) impulse response of a linear model. We will come back to this topic later (Sec. IV A) and present a brief description. As a result of the application of this theorem, the indirect method of determining the impulse response leads to an interpretation in terms of an equivalent linear system even though the actual system (the cochlea) is nonlinear and the degree of nonlinearity is randomly fluctuating in time. In the present paper we describe properties of direct and indirect impulse response functions as they have been measured mechanically in the basal turn of the guinea-pig cochlea. We will concentrate on low-level phenomena and, through the use of the aforementioned EQ-NL theorem, interpret what happens at higher levels of stimulation where the cochlea is nonlinear.

II. DATA FROM CURRENT EXPERIMENTS

A. Experimental method

Recordings were made of the movements of the basilar membrane (BM) in the cochlea of anesthetized guinea pigs. This study was consistent with NIH guidelines for humane treatment of animals and was reviewed and approved by the University of Michigan Committee on Use and Care of Animals. Pigmented guinea pigs were surgically prepared to allow visual access to the scala tympani surface of the basilar membrane in the basal or first turn of the cochlea. A ventral and postauricular dissection exposed the cochlea leaving the middle ear ossicles and tympanic membrane intact. A small opening was made in the bony wall of the cochlea of the scala tympani in the first turn and gold-coated glass microbeads (diameter smaller than 20 μm) were deposited onto the scala tympani surface of the basilar membrane. The beam of a laser Doppler vibrometer (OFV 1102) was directed through a microscope and focused onto a glass bead, the movement of which registers the basilar membrane motion (Nuttall

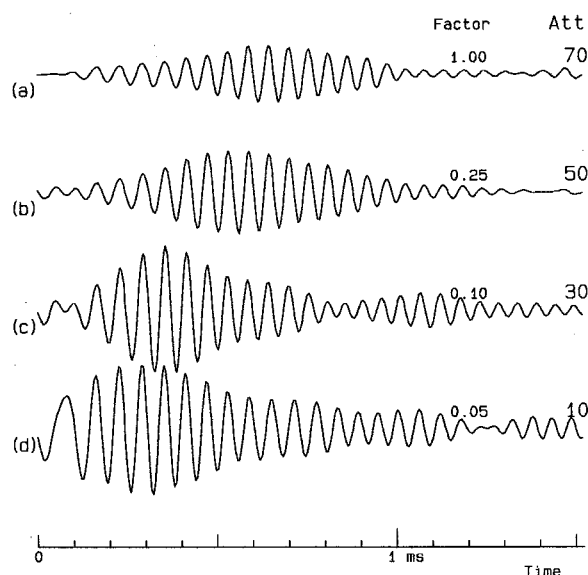


FIG. 1. Waveforms of impulse responses of the basilar membrane. Experiment: 0620 (guinea pig), file: #023. Impulse duration: $24 \mu\text{s}$ (5 sample periods). Intensity levels indicated by attenuation value ("Att") with respect to 100 dB peak SPL. Response waveform shown over 1.52 ms. Signal values are plotted multiplied by the "Factor" indicated in the figure. Signals are corrected for the stapes response. (a) Att=70, a weak stimulus; (b) Att=50, some deviant behavior in the beginning; (c) Att=30, the initial part of the waveform shows a more clearly different behavior; (d) Att=10, initial oscillations are widely different from later ones.

et al., 1990, 1991). During recording a thin glass plate over the hole served to correct for optical distortion and to minimize vibration of the fluid surface.

Signal generation and data acquisition were done with our own custom-designed programming system, which operates with a clock frequency of 208 kHz. The system and our method of data acquisition are briefly described in Appendix B.

The total duration of an experiment with an "intact" (i.e., optimally functioning) cochlea varied between 4 and 13 h. After the experiment, the animal was sacrificed and a number of records was taken from the "dead" cochlea. Finally, the laser beam was directed at the stapes, and frequency response and impulse response were carefully measured again from a glass bead (diameter $100 \mu\text{m}$) placed on the stapes footplate. In the present paper all our BM response data are corrected for the stapes response. This implies that our data are corrected for transducer response and middle ear transmission. Omitting this correction does not alter the nature of our results as described here because the stapes response is wide-band.

The programming language for experiments and analysis is Turbo Pascal[®] (version 7.0). Programming of data handling, experiments and analysis has mainly been done by the first author; the actual experiments have been carried out in the laboratory of the second author. Final analysis of the data was done in close collaboration.

B. Typical results—Direct impulse response

Figure 1 shows four direct impulse responses (BM velocity records), obtained at various stimulus intensities.

These data are obtained from a cochlea that is (as far as we can judge) functioning optimally. For weak stimuli the best frequency (BF) of the cochlear location studied is 17.9 kHz. The stimulus signal was a pulse, 5 sample periods long (approximately $24 \mu\text{s}$), which was repeated with a period of approximately 19.7 ms. It should be noted that in our figures stimulus intensity is always denoted by its complement, the attenuation ("Att") with respect to the level of 100 dB. At 0-dB attenuation the sound pressure during the pulse is equal to the peak pressure of a sinusoidal signal of approximately 105 dB SPL. Some more details of data processing are given in Appendix B.

The figure shows the impulse response over most of its main lobe of oscillations. The response continues after the period shown, and has, in some of our animals tested, a few later lobes. We do not discuss these later lobes here. By choosing different multiplication factors ("Factor" in the figure) we show the waveforms as having comparable amplitudes. One of the curves (panel a) has been recorded at such a low intensity that we may expect that nonlinear effects are minimal. In the other three (b, c, and d) the stimulus intensity increased in steps of 20 dB, and cochlear nonlinear effects are observable. With increasing intensity the impulse response appears to start earlier and earlier, and has a character quite different from the response at low stimulation levels. We also observe that the response appears to start abruptly at time zero. This phenomenon is due to the windowing that we applied in the frequency domain which is the same for all records (and rather severe, see Appendix B). At the higher stimulus intensities the bandwidth of the response is larger and the windowing has more effect. We consider as more important that *in the initial part of the response the frequency of oscillation is lower than in the remaining part*. A simple explanation would be one in terms of a temporally varying degree of nonlinearity (see Sec. I). To study this feature more accurately, we have applied a special method to isolate the "instantaneous frequency of oscillation"; the technique used is based on the concept of the "analytic signal" and is briefly described in Appendix C. Figure 2 shows three of the four waveforms from Fig. 1, with plots of the instantaneous frequency (IF) added. The IF scale is linear but covers only a relatively small frequency region, namely, from 15 to 20 kHz. The IF is only drawn where the amplitude of the impulse response is less than 12 dB below the maximum amplitude, this is done because noise in the data prevents us from deriving a reliable IF in the periods with much smaller amplitudes. The same sensitivity to noise of the IF method shows up in the form of irregularities in the IF curves (in this as well as in later figures). The reader's attention is directed at the gradual course of the IF. It is seen that in the high-level record (c), the oscillations start with a low frequency and the IF increases throughout the main part of the response. What is particularly striking: the same is the case in the *low-level* recording (a). The IF increase appears to continue until after the amplitude of the waveform starts to decrease. In this, the low stimulus-level case, we have no reason to assume that the cochlea would react in a nonlinear way during the initial part of the response. We will refer to this phenomenon, the gradual increase of the IF during the

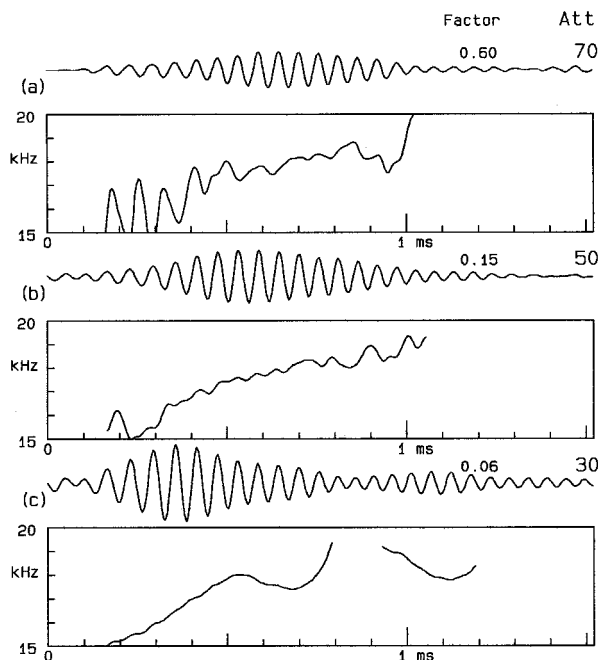


FIG. 2. Waveforms and instantaneous frequencies (IFs) of impulse responses of the basilar membrane. Experimental conditions the same as in Fig. 1. Below each waveform is shown the IF curve enclosed in a rectangle. The linear IF scale goes from 15 to 20 kHz. The time scale is indicated on the abscissa. The IF is only drawn where the amplitude is less than 12 dB below the maximum amplitude. Panels (a) to (c) as in Fig. 1; panel (d) is omitted.

main part of the impulse-response waveform, as the “glide,” and note that it appears to be present over a wide range of stimulus intensities. Figure 3 shows other examples of direct impulse responses and records of the IF for low-level

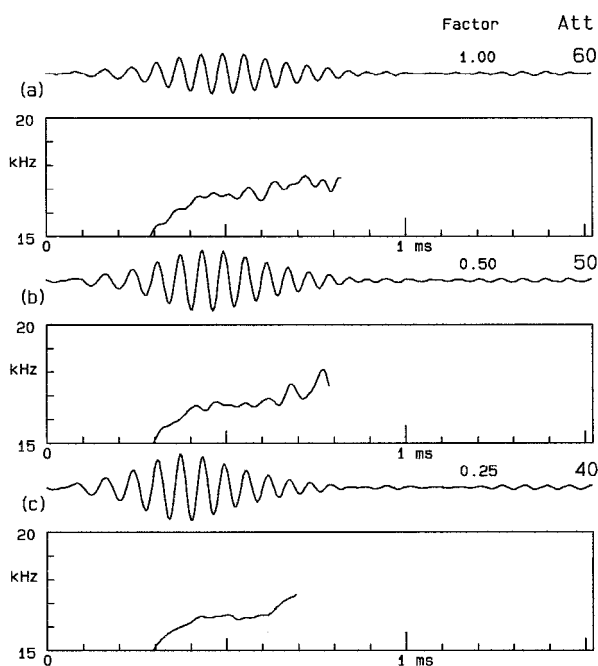


FIG. 3. Waveforms and instantaneous frequencies (IFs) of impulse responses of the basilar membrane; low-level case. Experiment: 0621 (guinea pig), file: #009. General layout as in Fig. 2.

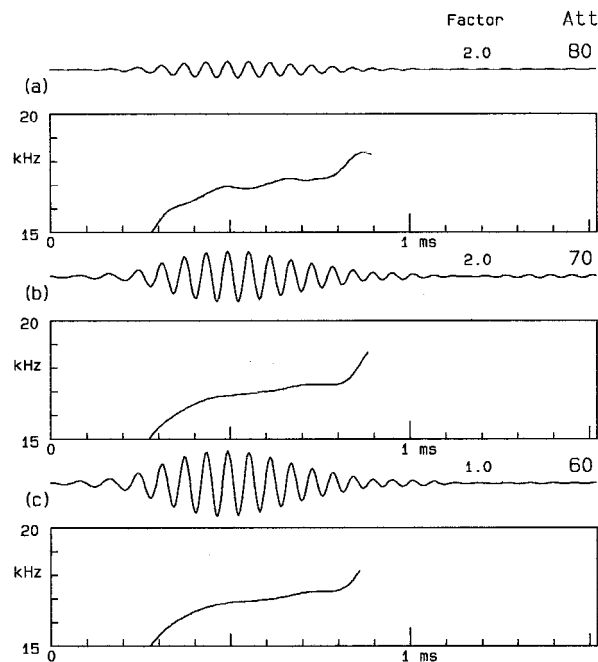


FIG. 4. Cross-correlation functions (ccfs) and instantaneous frequencies (IFs) of the basilar membrane. Experiment: 0621 (guinea pig), file: #001, optimally functioning cochlea, very weak stimuli. Stimulus: a band of noise two octaves wide, centered at the BF (18 kHz). “Att” values indicate attenuation in dB with respect to 95 dB over-all SPL of the noise stimulus. General layout as in Fig. 2.

stimuli, for another animal. In all three recordings we find a glide to be present, and the frequency of oscillations continues to increase throughout the main lobe of the response.

C. Typical results—Indirect impulse response (ccf)

We have complemented our findings with measurements of the indirect impulse response. That is, we applied a stimulus consisting of stationary wide-band noise, centered at the best frequency of the location studied, and computed the cross-correlation function (ccf) between stimulus and response signals. We believe that at low levels the cochlea operates as a linear system. Then, the ccf measured with wide-band noise should be indistinguishable from the system’s impulse response (Appendix A), so it comes as no surprise that ccfs show the glide phenomenon too. Figure 4 shows three examples, from records measured at very low intensities in the same animal as used for Fig. 3. Figure 5 shows similar results, from a different animal, but (in one record) at a still lower level of stimulation. In all panels of this figure we observe the “glide” clearly.

From the figures we see that, when the cochlea is strongly stimulated, the impulse response appears to start earlier than when the stimulus is weak. It may be asked whether the glide that we observe with weak stimuli can be considered as an *extrapolation* of the glide in the response with strong stimuli (or, even of the dead cochlea). To study this subject deeper, we present Fig. 6 showing results over a very wide range (80 dB!) of intensity levels. Note that the IF frequency scale has been made wider, from 10 to 20 kHz. In all panels the IF function for the 70-dB attenuation response is shown by a dashed line, and this serves as a guide. In

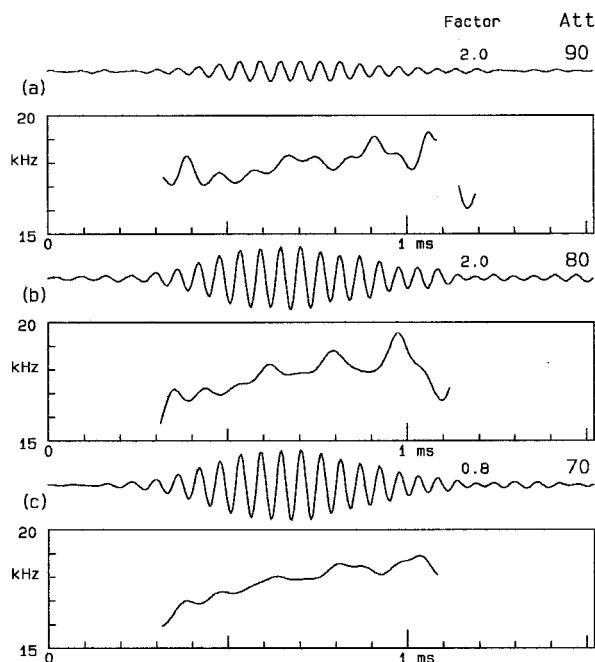


FIG. 5. Cross-correlation functions (ccfs) and instantaneous frequencies (IFs) of the basilar membrane. Experiment: 0620 (guinea pig), file: #019, optimally functioning cochlea, very weak stimuli. Layout as Fig. 4.

panel (c), the high-level stimulus case, the IF is seen to rise from 10 kHz, almost immediately after the oscillations start. The IF curve is soon interrupted because the amplitude of the waveform starts to decrease early in this case. In the other panels, referring to weaker stimuli, the IF curve starts later because the waveform amplitudes rise only slowly to their

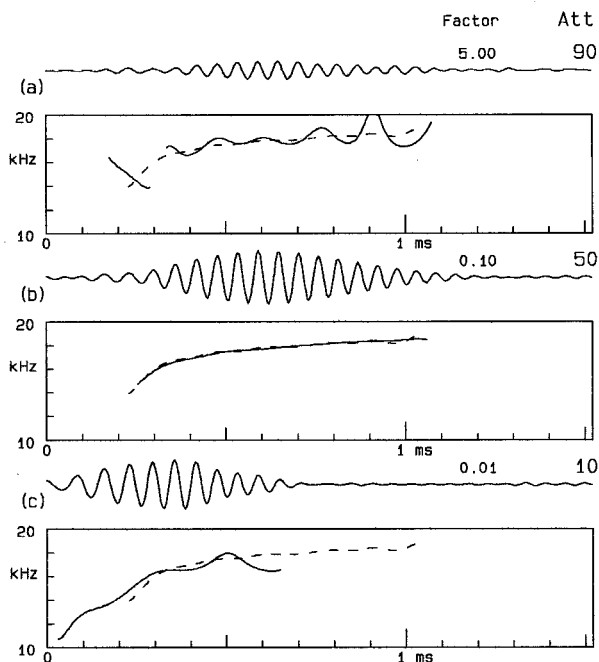


FIG. 6. Cross-correlation functions (ccfs) and instantaneous frequencies (IFs) of the basilar membrane. Experiment: 0628 (guinea pig), file: #035. This figure illustrates the effect of stimulus intensity over a very wide range. Layout as Fig. 4. The dashed line shows the IF function for the 70-dB attenuation condition.

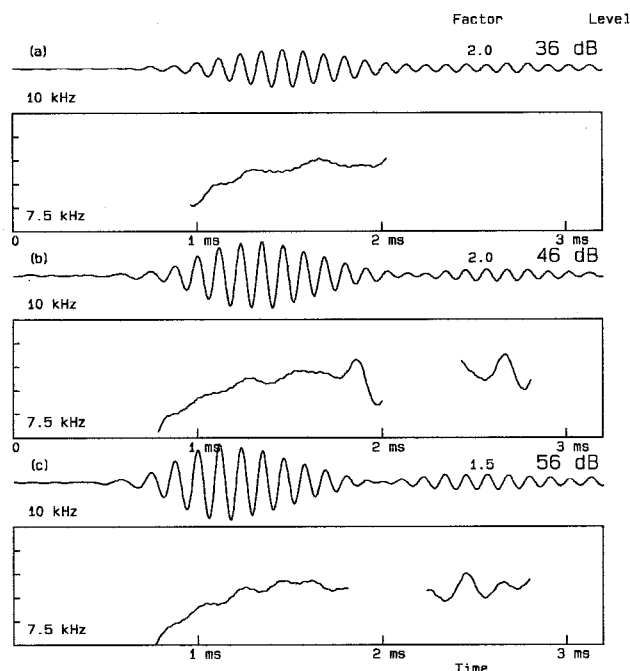


FIG. 7. Impulse response and IF curve for a location in the basal turn of the chinchilla cochlea (Courtesy Mario Ruggero). Layout as in Fig. 2, note that IF scale encompasses lower frequencies. The dB values are peak-SPL values, see the reference. At the lowest level the best frequency is 9.0 kHz.

maximum amplitude. The rise of the IF is continued to beyond the time of the maximum amplitude. In fact, the rise of the IF continues over nearly the entire time that the response is observable, its slope decreasing slowly. Closely related to the similarity of the IF functions is the property that the phase of the oscillations does not depend much on stimulus level. In point of fact, the individual lobes of the ccf waveform, considered for different stimulus levels, line up rather neatly. This is a most interesting and intriguing subject by itself which will be a topic of study in later work.

III. DATA FROM OTHER EXPERIMENTS

A. Mechanical measurements

The authors have been provided with data from experiments performed by colleagues, and these data have been processed with the aim of trying to discover similar trends in equivalent impulse responses. In one set of data, direct impulse responses in the basal turn of the chinchilla cochlea (Ruggero *et al.*, 1992, courtesy Mario Ruggero), this was particularly easy. We applied high-pass filtering with a cut-off frequency of 3 kHz, and low-pass filtering with a cut-off frequency of 10 kHz to a subset of his data. Figure 7 shows a typical example of low-level impulse responses. It is clear that a similar type of “glide” is present in the response but note that the location of measurement has a best frequency (BF) of 9.0 kHz, *about one octave lower* than in our earlier figures (for the guinea pig).

We have also processed data from mechanical frequency-response measurements in the basal turn of the guinea pig from another laboratory (Sellick *et al.*, 1983). These data were interpolated and converted to reasonably

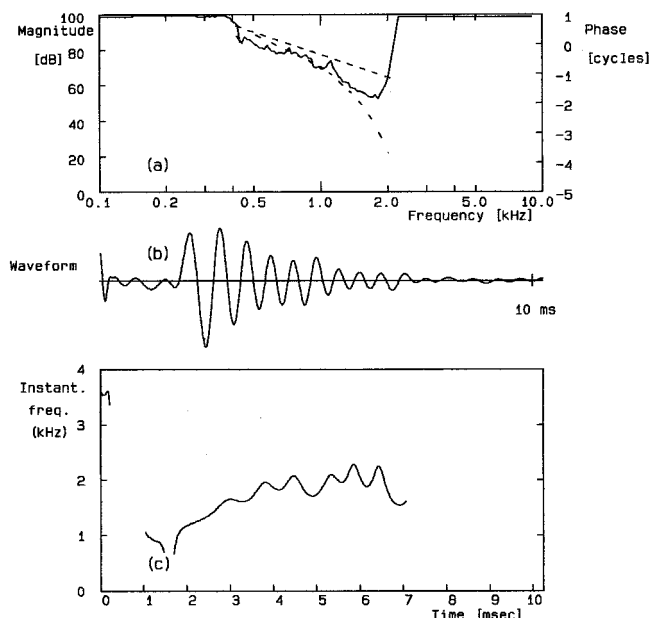


FIG. 8. Frequency threshold curve for neural response (cat data, courtesy Jont Allen/Paul Fahey) converted into impulse response. Unit: 9101, CF: 1.76 kHz, clock period used: 20 μ s, corrected for a delay of 2.0 ms. Panel (a): intensity and phase versus frequency. Solid line: frequency-threshold curve. The straight dashed straight line indicates the trajectory along which the phase data are taken. The curved dashed line shows the phase (i.e., the average phase of firing). Panel (b): reconstructed impulse response. Panel (c): instantaneous frequency versus time. Note: ordinate scale is linear and is centered near the CF of the unit.

smooth functions of frequency by a procedure described by Diependaal *et al.* (1987a). A simple Fourier transform was then applied to obtain a time response. The response functions were put at our disposition (courtesy Rob Diependaal) and processed. The results for nearly all of the data were entirely comparable to the ones we presented in Figs. 1 to 3, and thus we refrain from showing results.

B. Neural data

Data for mechanical BM responses in regions of lower frequencies would be most useful to study. However, suitable data are not available in mechanical form, therefore, we have to resort to other sources. In particular, we can draw information from recordings of neural activity. The first example comes from neural “revcor functions” (reverse-correlation functions). The data were collected in the first author’s laboratory by A. J. Breed (other, related, results of this work have been reported in Breed *et al.*, 1992, and de Boer and Breed, 1993). The experimental animal was the Mongolian gerbil, and recordings were made of single fibers of the auditory nerve. The findings corresponded fully to the results reported by Møller (1983), and we may conclude that the glide phenomenon persists at least down to the frequency range of 3 kHz.

Other neural data come from experiments (Allen, 1983) on frequency threshold curves (tuning curves) and response phase. The data were put at the authors’ disposal (courtesy Jont Allen and Paul Fahey), in the form of amplitude and phase records, and converted into equivalent impulse responses. Characteristic frequencies (CFs) for these data var-

ied from below 1 kHz to over 4 kHz (at higher frequencies no reliable phase data could be obtained). Appendix D gives some details about data processing.

In the interest of brevity we present only one figure. Figure 8 shows results for a primary fiber with a rather low CF (1.76 kHz). Panel (a) shows threshold level versus frequency, in the form of a solid line, and the phase as a curved dashed line. Phase was measured after the frequency threshold curve was determined; the straight dashed line shows the course of level versus frequency used in measuring the phase. Note that the phase data are taken at levels near and just above the threshold of the neural response. Panel (b) of the figure shows the reconstructed impulse response, and panel (c) the IF. The IF scale is again linear; it is centered at the CF but it encompasses the full range from zero. Again, we observe a clear glide.

IV. INTERPRETATION AND EVALUATION

A. Interpretation, the EQ-NL theorem

In the case of the direct impulse response a possible interpretation of the glide at high levels involves the notion that the degree of cochlear nonlinearity varies with time (see Sec. I). The same interpretation of a temporally varying degree of nonlinearity might be used in the case of the indirect impulse response (measured as a ccf) but the variations are random. However, this does not lead to more insight because the variations are difficult to predict quantitatively. There is another, and possibly better, interpretation. The word “interpretation” is used here in the sense that we consider a model of the cochlea and that we explain the observed phenomena in terms of phenomena occurring in that model. Let us consider a cochlear model that is linear as far as the fluid and the basilar-membrane proper are concerned, but that contains outer hair cells (OHCs) which act as nonlinear and compressive transducers. Without the contribution of the OHCs the response amplitude at a certain location x has a peak at a certain frequency f_0 . The notion “frequency selectivity” expresses the fact that the response (at the same place and for a constant stimulus level) is smaller at frequencies away from f_0 . At low stimulation levels the OHCs function to enhance frequency selectivity, i.e., to make the response variations around the peak frequency f_0 larger. OHCs are modeled to do this by producing a force (or pressure) that is an instantaneous function of the displacement of their stereocilia. This type of model is usually called “active” but a better term is “locally active” because at each frequency the response-enhancing activity of the OHCs is only needed over a limited region of the BM, basal to the location of the response peak (see de Boer, 1991). With strong stimulation the OHCs tend to become saturated and thus will contribute less to frequency selectivity. Call this nonlinear model M_1 .

For this type of nonlinear model of the cochlea a theorem has been derived, called the EQ-NL theorem (de Boer, 1997), which describes a property of the input–output ccf for wide-band signals. Explicitly, the theorem states that *the ccf for the nonlinear model M_1 is equal to the ccf for a comparison model M_2 that has exactly the same structure and the same parameters with the exception that it contains linear*

OHC transducers. The second part of the theorem states that in the comparison model M_2 the efficiency of all OHCs is reduced by a constant factor, which is the same for all OHCs.

Very briefly, the idea behind the theorem is the following. Consider each outer hair cell (OHC) as a memoryless nonlinear device embedded in a linear system (all frequency dependence of hair-cell transduction being transported to the “outside world,” i.e., the linear parts of the model). With a noise signal as input, the average compression of signal components by the nonlinear device is the same for all frequencies. In this sense the OHC transducer can thus be represented by a linear transducer with a frequency-independent attenuation.

When the nonlinear model is stimulated with strong sounds, the transduction of the OHCs will saturate (partly), and the extra force or pressure produced by them is reduced in a relative sense. As a consequence, frequency selectivity deteriorates. The same reduced selectivity will be demonstrated by the linear comparison model M_2 and this will also show up in its impulse response. In terms of the EQ-NL theorem, the ccf of the nonlinear system M_1 is equal to the impulse response of the linear comparison system M_2 . The glide can thus always be interpreted as a property of a *linear* system, and this is true at all levels of stimulation used.

In an extreme case, where the OHCs are entirely put out of action, we would observe the impulse response of a model simulating the “passive” (i.e., nonactive) cochlea. It is known that the impulse response of such a passive model has an intrinsic frequency modulation, a glide. See for the case of a passive long-wave model de Boer (1980), Figure 6.9. Let us now return from the passive case to the case of the normally functioning, active, cochlea, and do this in a gradual manner. With increasing degree of activity, the OHCs come into play to improve frequency selectivity more and more. Apparently, *this process occurs in such a way that the glide character is maintained.* In other terms, the impulse response is not swamped by the newly occurring, sharply frequency-selective, action of the OHCs. Nor is the new OHC action so purely that of a simple resonator that it dominates the response in all respects (a single resonator does not show a glide). What is even more surprising is that the course of the IF is nearly independent of stimulus level. In terms of the comparison model M_2 , the IF function is nearly independent of the efficiency of the OHCs.

B. Two locally-active cochlear models

We will next illustrate to which degree the glide is demonstrated by the response of two current locally active cochlear models. Both models are linear, and operate with the long-wave approximation. If the models were made nonlinear by assuming that OHCs are nonlinear transducers, both models would belong to the class of models to which the EQ-NL theorem (de Boer, 1997) applies.

We first choose a member of the class of *classical* models, a class characterized by the assumption that mechanical longitudinal coupling in the BM is absent and that longitudinal coupling occurs only via the cochlear fluid (cf. Viergever, 1986). In the Neely–Kim model (Neely and Kim,

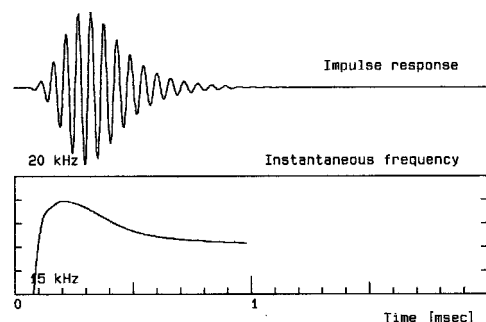


FIG. 9. Impulse response (upper panel) and IF curve (lower panel) of the Neely–Kim model. Model parameters as in Neely and Kim (1986); the effective channel height (channel area divided by BM width) was made equal to 0.4 (cm). Length of model divided into 500 sections. Time step: 2 μ s. OHC effectivity coefficient γ : 0.98. The IF curve is shown over the period where the signal is less than 40 dB below its maximal value. Average best frequency (BF) of impulse response at location selected: 18.1 kHz, BM resonance frequency at this location: 39.6 kHz.

1986) OHCs are assumed to contribute to the sound pressure in the cochlear channels. In order to achieve realistic local activity each OHC is assumed to be stimulated by movements that undergo extra filtering. The filtering is due to a resonance assumed to be exhibited by the tectorial membrane (TM), and the local TM resonance frequency is a (more or less fixed) fraction of the local resonance frequency of the BM. The operation of the OHCs is controlled by a coefficient γ that can be chosen between 0 and 1; $\gamma=0$ yields a passive model (no OHC action) and $\gamma=1$ produces a model of which the response is well comparable with experimental findings at low stimulation levels in the intact cochlea. The Neely–Kim model (note that it is formulated as a long-wave model) was set up with the parameters as published by Neely and Kim (1986), except that the height of the model was chosen as 0.4 (cm)—this was done in an effort to simulate that in a three-dimensional model the BM occupies only a fraction of the channel width. The impulse response was calculated with the robust solution method described by Diependaal *et al.* (1987b). Instead of an impulse, a doublet pulse (two consecutive time samples were given the values +1 and −1) was used as the stimulus signal; this serves to depress persisting low-frequency components in the response. Hence, it is the time derivative of the impulse response that has been computed. The OHC effectivity coefficient γ was made equal to 0.98 (higher values tend to cause computational instability). Figure 9 shows the result, for the location that has its best frequency (BF) equal to 18.1 kHz. The figure shows the impulse response and the course of the instantaneous frequency (IF) computed with the same method as before. It is seen that a very fast rise of the IF occurs in the initial period where the oscillations are too small to be visible (in this figure we can show the IF curve over a much wider range of amplitude values because the noise in the impulse response is very small). In the period where we observe the oscillations to rise in amplitude toward their maximum, the IF is seen to reach a maximum and to decrease—in other words, there is an *inverted* glide. In the final phase of the impulse response, the IF function tends to

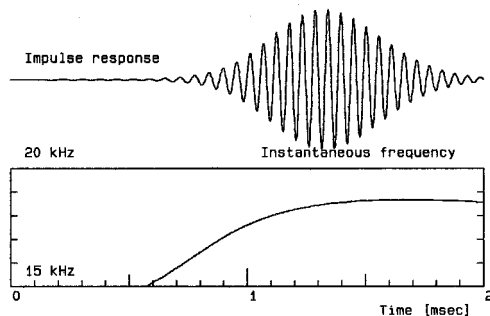


FIG. 10. Impulse response (upper panel) and IF curve (lower panel) of Geisler and Sang's feed-forward model (Geisler and Sang, 1995), TM resonance included. Model parameters as in the original paper, except for the effective channel height which was taken as 0.4 (cm) and the length of the model which was shortened to 1.5 (cm). Length of model divided into 250 sections. OHC effectivity coefficient γ : 0.45. The IF curve is shown over the period where the signal is more than 0.03 times its maximal value. Average best frequency (BF) of impulse response: 17.5 kHz.

settle at a constant value. We found similar behavior when the efficiency coefficient γ was given a lower value (for instance, 0.75, which corresponds to a loss of sensitivity in the response-peak region of approximately 30 dB). Note that, compared to the data, the group delay of this model is (much) too small, this can partly be attributed to our choice of the effective height. We conclude that the parameters of the Neely–Kim model are such that (at least for $\gamma > 0.75$) the impulse response does not show a glide as our (and others') data do. We do not know whether the model would perform better in this respect with different parameters.

The second model is the *nonclassical* model described by Geisler and Sang (1995) and conceived as a “feed-forward” model. In this model the OHCs receive their input from a location that is slightly more basal than the point where they are located. With this expedient the model shows many realistic properties. We used the form of the model with a secondary resonance in the TM (similar to that in the Neely–Kim model) to produce Fig. 10. Again, we chose the parameters as used by the authors but made the height equal to 0.4 (cm). We selected the efficiency factor so as to simulate the response of the intact cochlea at very low stimulation levels. However, we used a different (in fact, simpler) method for computing the impulse response because it is not known whether the Diependaal *et al.* (1987b) method remains robust when modified for nonclassical models. We computed the model response for 128 frequencies from 0 to 30 kHz, and stored the velocity response at a selected location of the BM in an array. Next we performed a Fourier transform on that array. The result is shown in Fig. 10, for the location that has 17.5 kHz as its best frequency (BF). Clearly, a very realistic glide is present, and we found it to change little in appearance when the efficiency of the OHCs in the model was reduced in order to simulate conditions where the cochlea is stimulated by stronger stimuli. Note that, in view of our data, the group delay of this model is too large (see comments in Geisler and Sang, 1995).

The glide phenomenon has been shown to be a fundamental property of data obtained from the cochlea. As such,

we propose that it becomes a standard to which models of the cochlea should be compared. It is interesting that the linear, passive long-wave model (de Boer, 1980, Figure 6.9) exhibits the glide while a later, locally active long-wave model, like the Neely and Kim (1986) model, fails the glide validation test (this paper, Fig. 9). Clearly, Geisler and Sang's (1995) model passes the proposed test (Fig. 10), although one cannot overlook the unreasonably long delay that is characteristic of this model. Both locally active models contain features that are not realistic, for instance, the assumed mass of the BM is too small. In view of our experimental and modeling results it appears that further research in the area of the glide could substantially enhance our understanding of the way the cochlea works.

ACKNOWLEDGMENTS

The authors wish to thank (in alphabetical order) Jont Allen, Arjan Breed, Rob Diependaal, Paul Fahey, Dan Geisler, Karl Grosh, Luc Johan Kanis, Dick Lyon, Bob Masta, Darren Miller and Mario Ruggero for their contributions to the present report. In addition, two anonymous reviewers have contributed much to the final presentation. This work has been supported by the Netherlands Foundation for Scientific Research (NWO), No. SLW 01.011, and by NIH, No. NIDCD-DC-00141.

APPENDIX A: AUTO- AND CROSS-CORRELATION FUNCTIONS

Consider a linear and time-invariant system H with impulse response $h(t)$ and apply an input signal $x(t)$ in the form of a white-noise signal with a power of unity. The *cross-correlation function* (abbreviated: ccf) $\varphi(\tau)$ between input signal $x(t)$ and output signal $y(t)$ is defined as

$$\varphi(\tau) = \mathcal{E}\{x(t+\tau)y(t)\}, \quad (\text{A1})$$

where the symbol $\mathcal{E}\{\cdot\}$ stands for “ensemble average” (or “expectation value”). This means an average over the ensemble of $x(t)$ signals having the same (given) probability distribution and the associated response signals $y(t)$. Instead of this procedure, averaging is usually carried out over time. Denoting the averaging time by T , Eq. (A1) then reads:

$$\varphi(\tau) = \int_0^T x(t+\tau)y(t)(dt/T) \quad (\text{for } T \rightarrow \infty). \quad (\text{A2})$$

When the signals $x(t)$ and $y(t)$ refer to the same signal, the ccf is known as the *auto-correlation function* (acf).

For the system H under consideration the following relation holds between the ccf and the impulse response:

$$\varphi(\tau) = h(-\tau). \quad (\text{A3})$$

In this paper Eq. (A3) is used to relate direct and indirect impulse responses to one another. This relation can easily be derived by using the convolution relation between $x(t)$ and $y(t)$:

$$y(t) = \int_0^\infty x(t-\tau)h(\tau)d\tau. \quad (\text{A4})$$

Relation (A3) is strictly true only for stimulation with white noise having unity power. When the input signal is not white noise, Eq. (A3) receives a correction including the acf of $x(t)$. However, when the bandwidth of the noise is considerably larger than that associated with $h(t)$, the correction is unimportant. This is the case for the low-level records used in the present paper. It is not entirely true for the high-level records and this results in visible distortion of the ccfs.

APPENDIX B: PROCESSING MECHANICAL RESPONSES

Our custom-designed programming system utilizes A/D and D/A converters located on PC boards (Computer Boards, type CIO-DAS 16/330i and cSBX-DDA4, respectively). The computer used was a Gateway 2000 (Intel 486, 50 MHz). Both D/A and D/A converters have a resolution of 12 bits, and operate with a clock frequency of 208 kHz. The stimulus signal consists of either pulses or a pseudo-random noise signal; one full period of the waveform is stored in a 4096-word array in memory (the “input-signal array”) and is sequentially read out to the D/A converter. The array is used as a *circular buffer*: When the end of the array is reached, the next sample is taken from the beginning, and so on. Consequently, the periods of the stimulus are presented consecutively, without any gaps or irregularities. Synchronously with the D/A conversions the signal from the laser velocimeter’s output amplifier is sampled by the A/D converter. The resulting numbers are stored in a temporary array. At regular intervals (synchronously with the periods of the stimulus) the contents of this array is added to the contents of another array, called the “output-signal array,” and a count (N) is maintained of the number of additions. In this way the output-signal array will, in due course, contain N times the average recorded BM velocity. Of course, this array is also interpreted as a circular buffer. In this paper we speak of “the response” as the average BM velocity recorded in this way, averaged after many (more than 1000) repetitions.

The D/A and A/D conversions occur concurrently and synchronously, both at a rate of 208 k samples per second. Since the arrays are 4096 samples long the duration of one period is 19.7 msec and the frequency components of the signal generated will be spaced by approx. 50.8 Hz. In the experiments the number N of repetitions is equal to or larger than 1000, hence each record requires a measuring time of at least 20 seconds.

The signal generated by the D/A converter is passed via a computer-controlled attenuator (Wilsons PATT) and an amplifier to a condenser microphone (B&K type 4134, 0.5" diameter) serving as a loudspeaker. A small acoustic coupler ensures optimal coupling to the ear drum membrane of the animal under test. No attempt has been made to compensate for the inherent quadratic distortion of a condenser microphone used as a loudspeaker. We checked that the quadratic distortion of the loudspeaker was smaller than -40 dB with respect to the primary level at the highest stimulus levels used, and we consider this small enough so as not to influence the results. Cubic distortion was smaller.

Experimental data have been acquired in arrays of 4096 long-integer words but are converted into (single-precision)

real variables for analysis purposes. Cross-correlation functions (ccfs) are computed via the digital Fourier transform of Eq. (A1) with the sign of τ inverted:

$$\Phi(\omega) = \mathcal{E}\{X(\omega) * Y(\omega)\}, \quad (\text{B1})$$

where the asterisk denotes the complex conjugate. This procedure is justified by the fact that the sampling frequency is well above the Nyquist frequency. Before further processing, low- and high-frequency components are reduced in amplitude. Except where indicated otherwise, the high-pass cut-off frequency is 6 kHz, and the low-pass cut-off frequency 20 kHz; the asymptotic slopes are 48 dB per octave for the direct impulse response and 4 times higher for the indirect impulse response. For the figures in this paper based on our own data, a correction for the stapes response has been applied. This we did via division of the respective spectra. For the direct impulse response (Figs. 1 to 3) we used the impulse response of the stapes. For the indirect impulse response we compensated via the ccf of the stapes response for a band-limited stimulus signal. It is because in this case the stapes ccf is a band-limited signal, and we divided by its spectrum, that we had to use higher cut-off slopes.

APPENDIX C: INSTANTANEOUS FREQUENCY AND THE “ANALYTIC SIGNAL”

Consider a real signal $x(t)$ which is zero for $t < 0$, and of which the Fourier transform $X(\omega)$ exists. The real part of this Fourier transform is an even, and the imaginary part an odd function of ω . The signal $x(t)$ may be an impulse response or a ccf. Because $x(t)$ is one-sided, real and imaginary parts of $X(\omega)$ are Hilbert transforms of one another. Now construct a new, one-sided, spectrum $Z(\omega)$ from $X(\omega)$ obeying the following conditions:

$$\begin{aligned} Z(\omega) &= 2X(\omega), & \text{for } \omega > 0, \\ Z(\omega) &= 0, & \text{for } \omega < 0, \\ Z(\omega) &= X(\omega), & \text{for } \omega = 0. \end{aligned} \quad (\text{C1})$$

Consider the inverse Fourier transform $z(t)$ of $Z(\omega)$. This function will have real and imaginary parts that are Hilbert transforms of one another because $Z(\omega)$ is one-sided. The real part of $z(t)$ is equal to $x(t)$ since for positive ω its spectrum is the average of $Z(\omega)$ and $Z(-\omega)$. Hence, the imaginary part of $z(t)$, let us call it $x_{\text{Hil}}(t)$, is the Hilbert transform of $x(t)$. The complex signal $z(t)$ is known as the “analytic signal.” It can be written as

$$z(t) = x(t) + ix_{\text{Hil}}(t), \quad (\text{C2})$$

or, alternatively, as

$$z(t) = a(t) \cos[\varphi(t)], \quad (\text{C3})$$

where $a(t)$ is the magnitude of $z(t)$ and $\varphi(t)$ the phase. The derivative of $\varphi(t)$ with respect to time t is the instantaneous frequency (IF). The just-outlined procedure is quite general but useful applications exist only for “narrow-band” signals, where variations of the magnitude $[a(t)]$ and the IF $[d\varphi(t)/dt]$ with time are slow with respect to the average IF.

This definition provides the path to compute the IF from a given real function of time $x(t)$. First, determine the Fou-

rier transform, apply the transformation given by Eq. (C1), and construct the analytic signal $z(t)$. Then, determine the phase $\varphi(t)$ and compute its time derivative (which means one has to correct for $\pm 2\pi$ phase jumps). This recipe has been followed to arrive at the figures showing the IF as a function of time; the application of the technique to sampled data presents no special problem. Prior to the computation of the IF the signal $x(t)$ (impulse response or ccf) undergoes extra smoothing, see Appendix B.

APPENDIX D: PROCESSING OF TUNING-CURVE DATA

Tuning-curve data come in the form of records of the measurement of the frequency threshold curve and, separately, records of the phase. Both are functions of frequency but the frequencies used are not the same, nor do the frequency values always form a regular series. Therefore, it is necessary to interpolate the data. The raw amplitude data give the series of turning points in the threshold measurement procedure, level and frequency. These data are interpolated in a linear way on a decibel scale. The series of frequencies used in measuring the phase is more regular, and phase-data interpolation is simple (provided appropriate corrections in steps of 2π are applied). Phase data are corrected for a delay as given in the figure legend, that delay includes the synaptic delay and (part of the) traveling-wave delay.

From the amplitude and phase data an array (of length 1024) is filled with the corresponding complex numbers. Finally, the so-obtained frequency response is windowed: components with frequencies lower than 2 octaves below the characteristic frequency (CF) are diminished in amplitude, and the same is done with components more than one octave above the CF. A Fourier transform is all that is needed to obtain the equivalent impulse response.

- Aertsen, A. M. H. J., and Johannesma, P. I. M. (1980). "Spectro-temporal receptive fields of auditory neurons in the grassfrog. I. Characterization of tonal and natural stimuli," *Biol. Cybern.* **38**, 223–234.
- Aertsen, A. M. H. J., Johannesma, P. I. M., and Hermes, D. J. (1980). "Spectro-temporal receptive fields of auditory neurons in the grassfrog. II. Analysis of the stimulus-event relation for tonal stimuli," *Biol. Cybern.* **38**, 235–248.
- Aertsen, A. M. H. J., Olders, J. H. J., and Johannesma, P. I. M. (1981). "Spectro-temporal receptive fields of auditory neurons in the grassfrog.

- III. Analysis of the stimulus-event relation for natural stimuli," *Biol. Cybern.* **39**, 195–209.
- Allen, J. B. (1983). "Magnitude and phase-frequency response to single tones in the auditory nerve," *J. Acoust. Soc. Am.* **73**, 2071–2092.
- Boer, E. de (1980). "Auditory physics. Physical principles in hearing theory. I," *Phys. Rep.* **62**, 87–174.
- Boer, E. de (1991). "Auditory physics. Physical principles in hearing theory. III," *Phys. Rep.* **203**, 125–231.
- Boer, E. de (1997). "Connecting frequency selectivity and nonlinearity for models of the cochlea," *Aud. Neurosci.* (in press).
- Boer, E. de, and Breed, A. J. (1993). "Encoding of 'nothing' in the peripheral auditory pathway: Nervous activity associated with a spectral gap," in *Brain Theory—Spatio-Temporal Aspects of Brain Function*, edited by A. Aertsen (Elsevier, Amsterdam), pp. 67–89.
- Breed, A. J., Kanis, L. J., and Boer, E. de (1992). "Cochlear nonlinearity for complex stimuli," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 189–195.
- Diependaal, R. J., de Boer, E., and Viergever, M. A. (1987a). "Cochlear power flux as an indicator of mechanical activity," *J. Acoust. Soc. Am.* **81**, 184–186.
- Diependaal, R. J., Duifhuis, H., Hoogstraten, H. W., and Viergever, M. A. (1987b). "Numerical methods for solving one-dimensional cochlear models in the time domain," *J. Acoust. Soc. Am.* **82**, 1655–1666.
- Geisler, C. D., and Sang, C. (1995). "A cochlear model using feed-forward outer-hair-cell forces," *Hearing Res.* **86**, 132–146.
- Möller, A. R. (1983). *Auditory Physiology* (Academic, New York), pp. 220ff and Fig. 3.13.
- Neely, S. T., and Kim, D. O. (1986). "A model for active elements in cochlear biomechanics," *J. Acoust. Soc. Am.* **79**, 1472–1480.
- Nuttall, A. L., Dolan, D. F., and Avinash, G. (1990). "Measurements of basilar membrane tuning and distortion with laser doppler velocimetry," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, Berlin), pp. 288–295.
- Nuttall, A. L., Dolan, D. F., and Avinash, G. (1991). "Laser Doppler velocimetry of basilar membrane vibration," *Hearing Res.* **51**, 203–214.
- 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, L. Demany, and K. Horner (Pergamon, London), pp. 85–91.
- Sachs, M. B., and Young, E. D. (1979). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," *J. Acoust. Soc. Am.* **66**, 470–479.
- Sachs, M. B., and Young, E. D. (1980). "Effects of nonlinearities on speech encoding in the auditory nerve," *J. Acoust. Soc. Am.* **68**, 858–875.
- Sellick, P. M., Yates, G. K., and Patuzzi, R. (1983). "The influence of Mössbauer source size and position on phase and amplitude measurements of the guinea pig basilar membrane," *Hearing Res.* **10**, 101–108.
- Viergever, M. A. (1986). "Cochlear macromechanics—a review," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, Berlin), pp. 63–72.

Fluid–structure interaction of the stereocilia bundle in relation to mechanotransduction

D. E. Zetes^{a)} and C. R. Steele

Division of Applied Mechanics, Stanford University, Stanford, California 94305

(Received 17 October 1996; revised 3 February 1997; accepted 4 February 1997)

Current hypotheses regarding mechanotransduction rely upon motion of the stereocilia relative to the apical surface of the hair cell. The viscosity of the surrounding endolymphatic fluid will, however, attenuate stereocilia motion at higher frequencies of excitation. To investigate stereocilia motion for physiologically reasonable deflections and frequencies of excitation, the fluid–structure interaction of the stereocilia bundle is considered analytically. Solutions in the frequency domain are determined for stereocilia bundle dimensions at several locations along the cochlear duct of the chinchilla. Results indicate that motion of the stereocilia is analogous to that of a low-pass filter. Comparison of these solutions with Greenwood's frequency-place map demonstrates that motion of the stereocilia bundle exists without substantial attenuation at least up to frequencies appropriate for the location of the corresponding hair cell along the cochlear duct. The variation in stereocilia morphology within the mammalian cochlea thus appears to provide a collection of low-pass mechanoreceptors, arranged in order of increasing corner frequency across the auditory spectrum.

© 1997 Acoustical Society of America. [S0001-4966(97)01906-1]

PACS numbers: 43.64.Kc, 43.40.At, 43.80.Gx [RDF]

INTRODUCTION

Deflection of the stereocilia bundle relative to the apical surface of the hair cell mediates mechanotransduction (Hudspeth and Corey, 1977; Shotwell *et al.*, 1981). Current hypotheses regarding gating mechanisms for operation of ion channels have in common a reliance upon motion of the stereocilia at auditory frequencies (Hudspeth, 1989; Hackney and Furness, 1995, for review). At present, however, it is not clear how motion of the fluid and surrounding structures influences displacement of the stereocilia. The viscosity of the endolymphatic fluid will attenuate the magnitude of stereocilia motion for higher frequencies of excitation (Steele *et al.*, 1993). Phase response and mode shapes of the stereocilia may also appear in the range of excitation frequencies. The fluid–structure interaction of the stereocilia bundle must therefore be investigated to determine its dynamic response for physiologically reasonable deflections and frequencies of excitation. Results can then be interpreted in relation to proposed models of mechanotransduction.

Previous investigators have considered mechanical analyses of the stereocilia bundle. Jacobs and Hudspeth (1990), Pickles (1993), and Geisler (1993) considered geometric models assuming rigid stereocilia and a specified relationship between the angular deflections of each stereocilium in a column. These models may be valid for physiologically reasonable deflections of the stereocilia bundle within the quasi-static range of excitation, although the nature of geometric constraints prevents these models from being extended to the dynamic case. Several investigators have also considered the response of the stereocilia

bundle using single degree of freedom lumped parameter models (Crawford and Fettiplace, 1985; Strelieff *et al.*, 1985; Allen, 1990; Freeman and Weiss, 1990; Assad *et al.*, 1992; Authier and Manley, 1995). Such single degree of freedom models consider movement of the stereocilia bundle as a whole, thereby neglecting hydrodynamic forces from motion of each stereocilium. These models are therefore restricted to analysis of a single mode shape. A more general multi-degree of freedom finite element model was developed by Duncan *et al.* (1993). Although motion of each stereocilium was considered, hydrodynamic forces from the surrounding endolymphatic fluid were not included.

Consideration of hydrodynamic forces is necessary for analysis of mechanotransduction and phase relationships at auditory frequencies of excitation. In the following analysis, linear, multi-degree of freedom equations of motion for the fluid–structure interaction of the stereocilia bundle including the extracellular linkages are derived and generalized for any finite number of stereocilia. Solutions in the frequency domain are considered for measured geometries of inner and outer hair cell stereocilia bundles from the chinchilla cochlea (Lim, 1980). No effort is made within this model to describe a particular ion channel or gating mechanism. Rather, only the motion of the stereocilia bundle relative to the apical surface of the hair cell is modeled to allow future quantitative investigation of hypotheses regarding mechanotransduction. These linear equations are valid until the amplitude of stereocilia motion causes the net force in the tip links to become compressive, at which point buckling may contribute a significant nonlinearity. An abbreviated form of this analysis appears in Zetes and Steele (1996), using stereocilia dimensions from the guinea pig cochlea measured in transmission electron microscopy (Zetes, 1995).

^{a)}Current address: Medtronic AneuRx, 10231 Bubb Road, Cupertino CA 95014.

I. ANALYTICAL DEVELOPMENT

Physiological deflections relevant for mechanotransduction are small (0.003° – 1° , Hudspeth, 1989) and parallel to the axis of morphological symmetry (Shotwell *et al.*, 1981). The apex of the cell also rotates only for large, physiologically unreasonable deflections of the stereocilia bundle (Strelioff and Flock, 1984). In the following analysis, linear deflections of the stereocilia in the excitatory/inhibitory plane are considered from the normal to the hair cell's apical surface. As observed by Flock *et al.* (1977) and Corey *et al.* (1989), when a stereocilium is loaded at its tip, the majority of elastic deformation occurs near its tapered base at its insertion into the apex of the hair cell. Each stereocilium shaft has thus been treated in this analysis as a rigid body with motion about its basal attachment resisted by an apparent rotational spring. This equivalent system is derived from Euler–Bernoulli beam theory for a load applied at the tip of a beam of varying circular cross section similar to that of a stereocilium. Finally, the linkages connecting the stereocilia are treated as linearly elastic elements, with spring constants derived from one-dimensional elasticity. For clarity, details of the derivation are described for a single column of three stereocilia and later generalized for n stereocilia in a given column.

A. System description and nomenclature

A column of three stereocilia as considered in this analysis is illustrated in Fig. 1. The angles between the long axes of the tall, middle, and short stereocilia with the normal to the cuticular plate are described by q_1 , q_2 , and q_3 , respectively. The lengths of the tall, middle, and short stereocilia are similarly denoted by l_i ($i=1,2,3$). The diameter of the stereocilia is denoted by d , the separation between the stereocilia rootlets by s , the distance from the base of the tall stereocilium to the upper attachment of the first tip link (connecting the tall and middle stereocilia) by lt_1 , and the distance from the base of the middle stereocilium to the upper attachment of the second tip link (connecting the middle and short stereocilia) by lt_2 . The angles the first and second tip links make with the line parallel to the surface of the cuticular plate are described by q_{12} and q_{23} , respectively. To approximate the resultant force on the stereocilia from the sound-induced shearing motion of the reticular lamina relative to the tectorial membrane (Pickles, 1988, for review), a force F_1 , is applied at the tip of the tallest stereocilium.

B. Linear equations of motion

Linear equations of motion are of the general form

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{F}, \quad (1)$$

where the vector \mathbf{q} and its time derivatives (denoted by a superposed dot) are 3×1 vectors with components representing the angular orientation of each stereocilium in a column of three stereocilia;

$$\mathbf{q} = \begin{Bmatrix} q_1 \\ q_2 \\ q_3 \end{Bmatrix}, \quad \dot{\mathbf{q}} = \begin{Bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{Bmatrix}, \quad \ddot{\mathbf{q}} = \begin{Bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \\ \ddot{q}_3 \end{Bmatrix}. \quad (2)$$

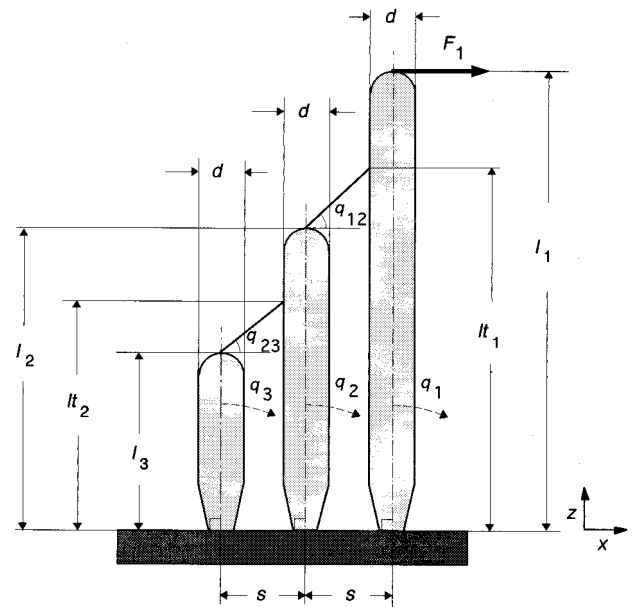


FIG. 1. Side view of a column of three stereocilia with the coordinate systems and notation used in this analysis. Angles q_1 , q_2 , and q_3 , respectively, describe the angle between the long axis of the tall, middle, and short stereocilia with the normal to the cuticular plate. Similarly, lengths l_i ($i=1,2,3$) denote the lengths of the tall, middle, and short stereocilia. The diameter of the stereocilia is denoted by d , the separation between the stereocilia rootlets by s , and the distance from the base of the stereocilia to attachment of the first tip link (connecting the middle and taller stereocilia), and the second tip link (connecting the short and middle stereocilia), by lt_1 and lt_2 , respectively. Angles q_{12} and q_{23} describe the angles the first and second tip links make with the line parallel to the surface of the cuticular plate, respectively.

In Eq. (1), \mathbf{M} is the mass matrix composed of the scalar components of inertia, \mathbf{C} is the damping matrix quantifying the effects of fluid viscosity, and \mathbf{K} is the stiffness matrix representing the coupled structural stiffness of the stereocilia bundle. These matrices are 3×3 for a column of three stereocilia, with their elements referred to in standard subscript notation

$$\mathbf{M} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}, \quad (3)$$

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix}.$$

Finally, the forcing function \mathbf{F} in Eq. (1), is a 3×1 vector representing the external forces applied to the stereocilia bundle. The loading case of a harmonic force applied to the tip of the tallest stereocilium is given by

$$\mathbf{F} = \begin{Bmatrix} F_1 l_1 \sin \omega t \\ 0 \\ 0 \end{Bmatrix}, \quad (4)$$

where F_1 is the magnitude of the force, ω is the frequency of excitation, and t represents time.

1. Stiffness contributions

During deflection of the stereocilia bundle, the stereocilia bend and all extracellular linkages undergo some elongation. However, using a geometric analysis, Geisler (1993) showed that the elongation of the lateral links is an order of magnitude smaller than the elongation of the tip links. Assuming that the lateral links have similar material properties to the tip links, the contribution of the lateral links to the structural stiffness of the stereocilia bundle is negligible. The major contributions to the elements of the stiffness matrix \mathbf{K} , therefore come from bending of the stereocilia and elongation of the tip links. The stiffness matrix can then be written as a matrix sum $\mathbf{K} = \mathbf{K}^{\text{cilia}} + \mathbf{K}^{\text{tip}}$, where $\mathbf{K}^{\text{cilia}}$ and \mathbf{K}^{tip} are contributions from the stereocilia and tip links, respectively.

The matrix $\mathbf{K}^{\text{cilia}}$ is diagonal and composed of the apparent rotational stiffnesses k_1 , k_2 , and k_3 , for the case of a load applied at the tip of the tall, middle, and short stereocilia, respectively,

$$\mathbf{K}^{\text{cilia}} = \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix}. \quad (5)$$

The form of these constants k_i ($i = 1, 2, 3$), in relation to the morphology and material properties of the stereocilia is detailed in Sec. C 1.

The matrix \mathbf{K}^{tip} is symmetric and tridiagonal

$$\mathbf{K}^{\text{tip}} = \frac{k_{\text{tip}}}{l_0} \times \begin{bmatrix} \lambda_{12}(\alpha l_1 + \beta_{12}r) & -\lambda_{12}\alpha l_2 & 0 \\ -\lambda_{12}\alpha l_2 & \eta_{12}\alpha l_2 + \lambda_{23}X & -\lambda_{23}\alpha l_2 \\ 0 & -\lambda_{23}\alpha l_2 & \eta_{23}\alpha l_3 \end{bmatrix}, \quad (6)$$

where k_{tip} is the linear stiffness of the tip link (also detailed in Sec. C 1), l_0 is the natural length of the tip link, s is the separation between stereocilia rootlets, and r is the radius of the stereocilia shaft. Geometric parameters are given by $\alpha = s - r$, $\beta_{12} = l_{t1} - l_{t2}$, $\lambda_{12} = l_{t1} \cos q_{12} - r \sin q_{12}$, $X = \alpha l_{t2} + \beta_{23}r$, and $\eta_{12} = l_{t2} \cos q_{12}$, and the angular orientation of the tip links is given by $q_{12} = \tan^{-1}((l_{t1} - l_{t2})/(s - r))$. Similar equations can be written for the second stereocilia pair to determine β_{23} , λ_{23} , and η_{23} .

2. Viscous contributions

Consideration of characteristic dimensions for physiological deflections and auditory frequencies demonstrates that flow between the stereocilia is predominantly linear and viscous (see the Appendix). Hydrodynamic contributions to the equations of motion can therefore be written in matrix form where in this analysis, the elements of the viscous matrix \mathbf{C} in Eq. (5), are derived from selected portions of lubrication theory. For the case of the stereocilia bundle, these approximate solutions assume that the dominant viscous effects occur in the region where the stereocilia are close to each other and their curvature is slowly varying. For fluid bearing surfaces of relatively small width and finite length,

the effect of a pressure gradient in the long direction is slowly varying in comparison to the short direction (Ocvirk, 1952). The dimensions of the stereocilia are such that the diameter is much smaller than the stereocilium length $d \ll l_i$ ($i = 1, 2, 3$). It is therefore reasonable and consistent to assume that the effect of pressure gradient in the long, or axial direction of the stereocilia is small.

The stereocilia bundle may be treated as having two fluid bearing surfaces, where the first bearing surface is between stereocilia pairs in a given column, and the second is between stereocilia pairs in a given row, as shown in Fig. 2. For a local coordinate system oriented as shown in Fig. 2 and originating on the excitatory face of a stereocilium surface, the flow field between stereocilia pairs in a given column is approximated by the familiar Poiseuille and Couette flows written in two dimensions

$$v \approx \frac{1}{2\mu} \frac{\partial p}{\partial y} (x^2 - h(y, z)x), \quad (7a)$$

$$w \approx \frac{Wx}{h(y, z)}, \quad (7b)$$

where v and w are the velocity profiles between stereocilia pairs in the y and z directions, respectively, W is the net velocity of stereocilia membranes in the z direction, $h(y, z)$ is the separation between the stereocilia shafts, μ is the viscosity of the fluid, and $\partial p / \partial y$ is the pressure gradient in the y direction. For clarity, this analysis considers the geometrical simplification of the stereocilia arranged in a cubic array. The separation between the cylindrical shafts of stereocilia pairs is approximated by

$$h(y, z) = h_0(1 + \gamma y^2 + \Delta qz), \quad |y| \leq \frac{d}{2}, \quad (8)$$

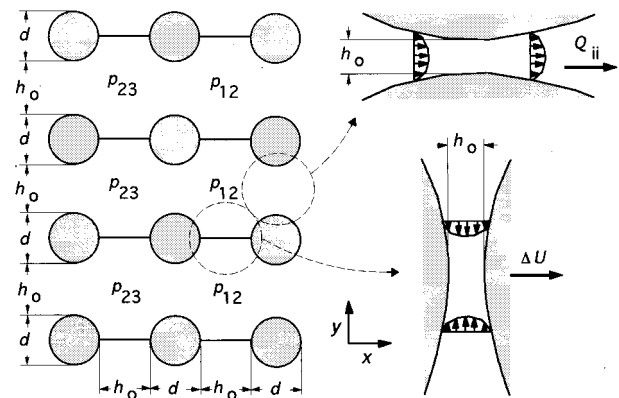


FIG. 2. Top view of the stereocilia bundle with the stereocilia rows and columns geometrically simplified for this analysis in the arrangement of a regular cubic array. The lines drawn connecting the stereocilia denote tip links connecting stereocilia columns. The smallest separation between the stereocilia shafts is denoted by h_0 , and the diameter by d . Relative pressures between stereocilia are denoted by p_{12} between the tall and middle stereocilia, and p_{23} between the middle and short stereocilia. Upper right corner illustrates assumed Poiseuille flow profile and net flow Q_{ii} , between stereocilia pairs in a given row. Lower right corner illustrates assumed Poiseuille flow profile and net velocity in the excitatory/inhibitory plane ΔU , between stereocilia pairs in a given column.

where $\gamma = 2/h_0 d$, h_0 is the smallest distance between the stereocilia shafts, Δq is the relative change in angular position between stereocilia pairs in a given column, and d is the diameter of the stereocilia. A more general stereocilia arrangement is considered in Zetes (1995).

For deflections in the excitatory/inhibitory plane, the stereocilia may also separate with a velocity component U in the x direction which does not affect Eqs. (8) and (9) and is considered when satisfying continuity. The relevant velocity components are given by

$$U_i = \dot{q}_i z, \quad (9a)$$

$$W_i = -\dot{q}_i (d/2) \quad (i=1,2,3), \quad (9b)$$

where the subscript i has been introduced to denote the linear velocities U_i and W_i , and angular velocities \dot{q}_i of each stereocilium ($i=1,2,3$).

Consideration of characteristic dimensions in a fashion similar to the Appendix demonstrates that hydrodynamic forces in the subtorcular space are an order of magnitude smaller than those between the stereocilia. They are therefore neglected in this analysis. Edge effects at either end of the "W" arrangement of stereocilia rows are also neglected.

a. Continuity. Continuity must be satisfied to ensure conservation of mass and compatibility of solutions between all stereocilia pairs. For small deflections of the stereocilia, flow in the long direction Eq. (7b), is predominantly a shearing of the fluid element with small net flow in comparison to flow in the short direction Eq. (7a). For this analysis, an additional approximation of planar flow is therefore introduced such that continuity is satisfied only in the plane parallel to the hair cell's apical surface. A more complete analysis considered in Zetes (1995) demonstrates that this approximation does not invalidate the previously described short bearing analysis nor significantly affect the numerical solutions described in Sec. II.

Global continuity is satisfied by considering control volumes at the intersection of stereocilia rows and columns

$$Q_{11} - Q_{22} = (U_1 - U_2)d, \quad Q_{22} - Q_{33} = (U_2 - U_3)d, \quad (10)$$

where Q_{11} , Q_{22} , and Q_{33} denote the net flow between stereocilia in the tall, middle, and short rows, respectively, and outward flow is defined as positive.

The net flows between the stereocilia rows Q_{ii} ($i=1,2,3$), are found from integration of the relevant velocity profile across the channel width and are given by

$$Q_{11} = \frac{2p_{12}h_0^3\gamma^{1/2}}{9\mu\pi}, \quad Q_{22} = \frac{2(p_{23}-p_{12})h_0^3\gamma^{1/2}}{9\mu\pi}, \quad (11)$$

$$Q_{33} = \frac{-2p_{23}h_0^3\gamma^{1/2}}{9\mu\pi},$$

where p_{12} and p_{23} are, respectively, the pressures between the tall and middle stereocilia, and the middle and short stereocilia, relative to ambient pressure external to the stereocilia bundle.

Combining Eqs. (9a), (10), and (11) and solving simultaneously, the relative pressures between the stereocilia are given by

$$p_{12} = \frac{3d\mu\pi(2\dot{q}_1 - \dot{q}_2 - \dot{q}_3)z}{2\gamma^{1/2}h_0^3}, \quad (12)$$

$$p_{23} = \frac{3d\mu\pi(\dot{q}_1 + \dot{q}_2 - 2\dot{q}_3)z}{2\gamma^{1/2}h_0^3}.$$

b. Traction and resultant moments. Contributions to the damping matrix \mathbf{C} in Eq. (3) are given by the resultant moments of the tractions acting on the stereocilia. The stresses within the fluid are determined from the constitutive relationship for an incompressible Newtonian fluid written in dyadic notation

$$\boldsymbol{\sigma} = -p\mathbf{I} + \mu(\nabla\mathbf{v} + (\nabla\mathbf{v})^T), \quad (13)$$

where $\boldsymbol{\sigma}$ is the stress tensor and \mathbf{I} is the unity tensor. Also in Eq. (13), the velocity profile \mathbf{v} is given by Eqs. (7)–(9) and similar equations for velocity profiles between stereocilia rows. The pressure distribution p is given by Eq. (12) and consideration of local continuity for each stereocilia pair.

The relevant tractions can be integrated over the stereocilia surfaces to obtain resultant moments about the bases of the stereocilia from the viscous and pressure forces. The resultant moment on the tall stereocilium for flow between stereocilia columns is given by

$$M_1^{\text{column}} = \int_0^{l_2} \int_{-\infty}^{\infty} \left(\sigma_{xx}|_{x=h} + \sigma_{xy}|_{x=h} \frac{\partial h}{\partial y} \right) z \, dy \, dz$$

$$+ \int_0^{l_2} \int_{-\infty}^{\infty} \sigma_{xz}|_{x=h} r \, dy \, dz, \quad (14)$$

where the inner limits of integration are infinite to allow the slowly varying approximation Eq. (8) to approach the far-field solution external to the journal bearing Eq. (12), and the outer limits of integration are to the height of the middle stereocilium, since as discussed previously, contributions from hydrodynamic forces in the subtorcular space are small.

Substituting for the necessary quantities and integrating,

$$M_1^{\text{column}} = \frac{\mu\pi l_2^3(1+4\gamma h_0^2)}{2\gamma^{3/2}h_0^3} (\dot{q}_1 - \dot{q}_2) + \frac{\mu\pi l_2 d^2}{4\gamma^{1/2}h_0} (\dot{q}_1 + \dot{q}_2). \quad (15)$$

Similarly, between stereocilia within the same row, the resultant moment on the tallest stereocilium is given by

$$M_1^{\text{row}} = \frac{2d\mu\pi l_2^3}{3\gamma^{1/2}h_0^2} (-2\dot{q}_1 + \dot{q}_2 + \dot{q}_3). \quad (16)$$

The total viscous moment acting on the tall stereocilium is found by superposition of the resultant moments in Eqs. (15) and (16). Resultant moments on other stereocilia are found in a similar fashion.

3. Inertia contributions

As shown in the Appendix, for physiological deflections and auditory frequencies, the hydrodynamic mass of the fluid is small and is therefore not included in the equations of motion. The mass of the stereocilia themselves may be included by considering a diagonal matrix composed of the

scalar mass moments of inertia of the tall, middle, and short stereocilium about their attachments to the cuticular plate. However, assuming the density of the stereocilia is of the same order of magnitude as water, the inertia terms of the stereocilia are also small in comparison to other coefficients in the equations of motion. They are therefore not described in detail here, since they do not contribute greatly for physiological frequencies of excitation.

C. Generalization for n stereocilia in a column

The matrix formulation of the equations of motion (1) can be generalized for stereocilia bundles containing more than three rows (i.e., in the bullfrog sacculle; Hudspeth, 1989) by considering a column of n stereocilia with coefficient matrices of dimensions $n \times n$. Referring to Eqs. (1) and (6), the nonzero elements of the upper half of the symmetric stiffness matrix \mathbf{K} , are given by

$$K_{ij} = \begin{cases} \frac{k_{\text{tip}}}{l_0} \lambda_{ii+1} (\alpha_{ii+1} l_i + \beta_{ii+1} r) + k_i, & i=j=1, \\ \frac{k_{\text{tip}}}{l_0} \eta_{ii+1} \alpha_{ii+1} l_i + \lambda_{ii+1} (\alpha_{ii+1} l_i + \beta_{ii+1} r) + k_i, & 1 < i < n, j=i, \\ -\frac{k_{\text{tip}}}{l_0} \lambda_{ii+1} \alpha_{ii+1}, & i < n, j=i+1, \\ \frac{k_{\text{tip}}}{l_0} \eta_{ii+1} \alpha_{ii+1} + k_i, & i=j=n, \end{cases} \quad (17)$$

where the tallest stereocilium is denoted by $i=1$ and the shortest stereocilium by $i=n$. The lengths of the stereocilia are denoted by l_i , the distance from the insertion of the i th stereocilia to the apical surface of the hair cell to the upper attachment of the tip link to the stereocilia shaft is denoted by l_{ti} , and geometrical parameters α_{ij} , β_{ij} , λ_{ij} , η_{ij} , and q_{ij} are found by analogy with those in Sec. B 1.

Similarly, referring to Eqs. (1), (15), and (16), the elements of the upper half of the symmetric damping matrix \mathbf{C} , are given by

$$C_{ij} = \frac{\mu \pi}{\gamma^{1/2} h_0} \begin{cases} \frac{l_2^3 (1 + 4 \gamma h_0^2)}{2 \gamma h_0^2} + \frac{l_2 d^2}{4} - \frac{4 d l_2^3}{3 h_0}, & i=j=1, \\ -\frac{l_{i+1}^3 (1 + 4 \gamma h_0^2)}{2 \gamma h_0^2} + \frac{l_{i+1} d^2}{4} + \frac{2 d l_{i+1}^3}{3 h_0}, & i < n-1, j=i+1, \\ \frac{2 d l_{i+1}^3}{3 h_0}, & i < n, j=i+2, \\ \frac{l_{i+1}^3 (1 + 4 \gamma h_0^2)}{2 \gamma h_0^2} + \frac{(l_{i+1} + l_i) d^2}{4} - \frac{2 d (l_{i+1}^3 + l_i^3)}{3 h_0}, & 1 < i < n, j=i, \\ -\frac{l_i^3 (1 + 4 \gamma h_0^2)}{2 \gamma h_0^2} + \frac{l_i d^2}{4} + \frac{2 d l_i^3}{3 h_0}, & i=n-1, j=n, \\ \frac{l_n^3 (1 + 4 \gamma h_0^2)}{2 \gamma h_0^2} + \frac{l_n d^2}{4} - \frac{4 d l_n^3}{3 h_0}, & i=j=n. \end{cases} \quad (18)$$

1. Generalization of stiffness constants

The stiffness constants appearing in Eqs. (5), (6), and (17) must be written in terms of material properties to accommodate stereocilia bundles of arbitrary dimensions. The apparent rotational spring constants k_i ($i=1,2,3,\dots,n$) can be written for a beam of varying cross section similar to that of a stereocilium using Euler–Bernoulli beam theory. For the case of an isotropic fixed-free beam with a circular cross-section, a conical taper at the fixed end, and a load F applied at the tip, as illustrated in Fig. 3, the apparent rotational spring constant is given by

$$k_{\text{rot}} = \frac{3 E_{\text{cilia}} \pi r_{\text{max}}^4 r_{\text{min}}^3 (l m + r_{\text{min}})}{4 l (r_{\text{max}}^4 + l m r_{\text{min}}^3 + r_{\text{min}}^4)}, \quad (19)$$

where k_{rot} is defined from moment equilibrium such that $F l = k_{\text{rot}} q$, where l is the length of the stereocilium, and q is the angle of rotation. Also in Eq. (19), r_{max} is the radius of the stereocilia in the cylindrical portion, r_{min} is the radius of the stereocilia at its insertion into the hair cell apex, l_{taper} is the length over which the change in radius occurs, $m = (r_{\text{max}} - r_{\text{min}})/l_{\text{taper}}$, and E_{cilia} is the Young's modulus of the stereocilia. The apparent rotational spring constant for each stereocilium k_i ($i=1,2,3,\dots,n$) is then given by Eq. (19) with l replaced by lengths l_i ($i=1,2,3,\dots,n$), respectively.

The linear stiffness of the tip links k_{tip} can be written in terms of material properties of the linkage constituent for a uniaxial elastic bar

$$k_{\text{tip}} = \frac{E_{\text{tip}} \pi r_{\text{tip}}^2}{l_0}, \quad (20)$$

where r_{tip} is the radius of the tip link, E_{tip} is the Young's modulus of the tip link, and l_0 is the natural length of the unstretched tip link.

D. Solution for frequency response

The frequency response of the stereocilia can be considered by a separation of variables solution, where the time dependence is oscillatory. The vector describing the angular deflections of the stereocilia is given by

$$\mathbf{q} = \text{Im}[\tilde{\mathbf{q}} e^{i \omega t}], \quad (21)$$

where ω is the frequency of the input force, $\tilde{\mathbf{q}}$ is a complex coefficient, and the imaginary part is taken to correspond to the phase of the input force as assumed in Eq. (3).

Substituting Eq. (21) into the linear equation of motion (1) gives

$$[-\mathbf{M} \omega^2 + i \mathbf{C} \omega + \mathbf{K}] \tilde{\mathbf{q}} = -i \tilde{\mathbf{F}}, \quad (22)$$

where $\tilde{\mathbf{F}}$ is defined similarly to $\tilde{\mathbf{q}}$ in Eq. (21), as the imaginary form of the forcing vector given in Eq. (4). For a column of n stereocilia, vectors $\tilde{\mathbf{q}}$ and $\tilde{\mathbf{F}}$ have dimensions $n \times 1$, and the coefficient matrix has dimensions $n \times n$.

Quantitative results for the amplitude and phase motion of each stereocilium as a function of the frequency of the input force are found from solution of Eq. (22). For interpretation of results in relation to mechanotransduction, a specific gating mechanism regarding tension in the tip links for

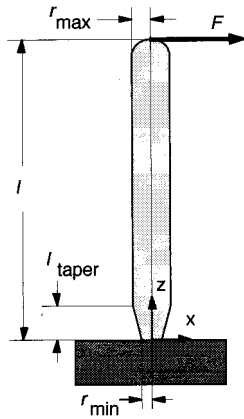


FIG. 3. A clamped-free beam with varying circular cross section similar to that of a stereocilium. A load F is applied at the tip of the stereocilium for derivation of an equivalent system from Euler-Bernoulli beam theory of a rigid stereocilium shaft with an apparent rotational spring constant k_{rot} at its base. The parameter l denotes the length of the stereocilia, r_{max} the radius of the stereocilia shaft, r_{min} the radius of the stereocilia at insertion into the hair cell's apical surface, and l_{taper} the length over which the change in diameter occurs.

operation of ion channels is considered (Pickles, 1988, for review). Within this analysis, the tension in the tip links connecting neighboring stereocilia is given by

$$T_{ij} = \frac{k_{tip}}{l_0} [(l_j r - s l t_i) q_i + l_j (s - r) q_j], \quad j = i + 1, \quad (23)$$

where the subscript ij denotes the sequential order of the stereocilia described previously, i.e., T_{12} is the tension in the tip link connecting the first and second stereocilia.

II. RESULTS

A. Amplitude solutions

The relationship between the tension in the tip links as a function of the frequency of the input force is shown in Fig. 4(a), where the tension in the tip links has been normalized by the magnitude of the applied load at the tip of the tallest stereocilium. To generate Fig. 4(a) dimensions for OHC1 from the basal region of the chinchilla cochlea were used (Lim 1980): $l_1 = 1.6 \mu\text{m}$, $l_2 = 1.3 \mu\text{m}$, $l_3 = 0.9 \mu\text{m}$, $d = 200 \text{ nm}$, $s = 400 \text{ nm}$, $l_{ti} = l_i + d$ ($i = 1, 2$), $r_{min} = d/4$, $l_{taperi} = l_i/4$ ($i = 1, 2, 3$), and $r_{tip} = 3 \text{ nm}$. The Young's modulus of both the tip links and stereocilia were $E_{tip} = E_{cilia} = 10^7 \text{ N/m}^2$, and the viscosity and density of water were used. As shown in Fig. 4(a) when a low-frequency force is applied to the tallest stereocilium, the normalized amplitude of the tension in the tip links is relatively large and constant. At higher frequencies of the input force the tension in the first tip link is attenuated in a fashion similar to a low-pass filter, while the tension in the second tip link demonstrates a modest "resonance" before attenuation. At very high frequencies, the viscous forces become so large that motion of the stereocilia bundle is prevented and a relatively small tension exists in the tip links. This behavior is consistently demonstrated in the first and second tip link for all stereocilia bundle geometries from the chinchilla cochlea for both inner and outer hair cells as reported in Lim (1980).

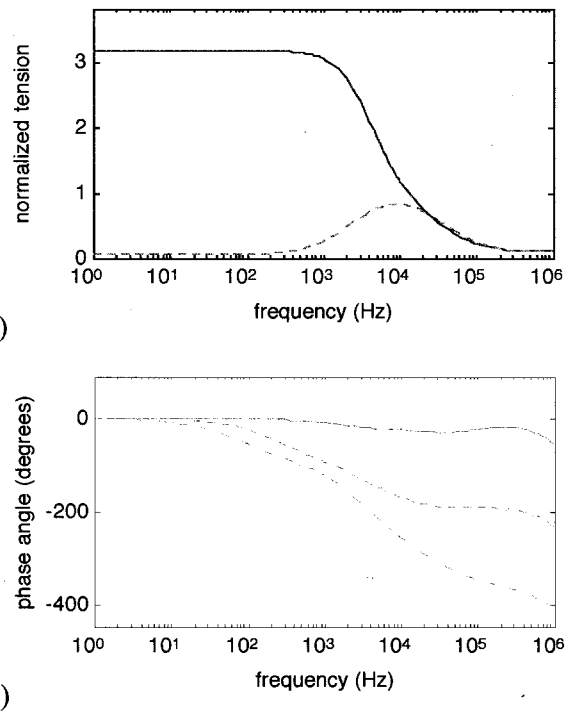


FIG. 4. (a) Frequency response of tension in the tip links normalized by the magnitude of the applied load at the tip of the tallest stereocilium. The response of the first tip link is similar to that of a low-pass filter, with amplitude attenuated for higher-frequency excitation. The tension in the second tip link demonstrates a modest "resonance" before attenuation. (— first tip link, --- second tip link). (b) Phase difference of angular displacement of stereocilia relative to the phase of the applied load. All stereocilia move in phase for quasi-static, low-frequency excitation. At higher frequencies, the stereocilia begin to exhibit phase response, approaching the viscous dominated limit of all stereocilia moving one-quarter cycle out of phase above the highest audible frequencies of the chinchilla. (— tall stereocilium, --- middle stereocilium, ... short stereocilium). In both figures, measured dimensions are for a stereocilia bundle from OHC1 in the basal region of the chinchilla cochlea (Lim, 1980), viscosity and density of water, and Young's modulus of the stereocilia and tip links $E_{cilia} = E_{tip} = 2 \times 10^7 \text{ Pa}$.

B. Phase solutions

The phase difference between the angular position of each stereocilium and the input force can also be determined from solution of Eq. (22). As shown in Fig. 4(b), for the same geometrical and material properties used in generation of Fig. 4(a), the stereocilia demonstrate continuously varying phase solutions. At low excitation frequencies, the stereocilia move in phase with each other and the input force. At higher frequencies of excitation, viscous effects contribute and the stereocilia begin to separate, such that the short stereocilium moves out of phase with the input force. At higher frequencies, the middle stereocilium also begins to move out of phase, and finally the tall stereocilium moves out of phase with the input force. Above the highest audible frequencies of the chinchilla, all of the stereocilia move toward the viscous dominated limit of a quarter cycle out of phase with the input force. As with the amplitude solutions described previously, similar behavior was observed for all stereocilia bundle geometries reported in Lim (1980).

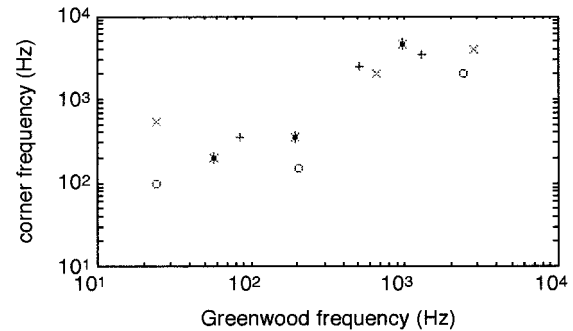
C. Dependence on stereocilia morphology

The mammalian cochlea is arranged tonotopically, with neural sensitivity to frequency varying exponentially along the cochlear duct. The morphology of the stereocilia bundle is also arranged tonotopically, with shorter stereocilia bundles in the higher frequency regions, and longer stereocilia bundles in the lower-frequency regions (see, e.g., Lim, 1980, and Strelioff and Flock, 1984). It is therefore interesting to investigate the relationship between stereocilia morphology and frequency response of the stereocilia bundle while holding other parameters, including the loading case, constant. In Fig. 5, the corner frequencies, defined as 3 decibels below the quasi-static solution from the current analysis, are shown using stereocilia lengths reported in Lim (1980) for all three rows of outer hair cells and one row of inner hair cells. These solutions are plotted versus the corresponding frequency by converting the length of the tallest stereocilium to the location along the cochlear duct (Lim, 1980) and this in turn to a frequency location using the frequency-place map for the chinchilla from Greenwood (1990).

In Fig. 5, it can be seen that the corner frequency of stereocilia motion is slightly above the Greenwood frequency location of the corresponding hair cell. These results demonstrate that motion of the stereocilia bundle relative to the apical surface of the hair cell exists without substantial attenuation at least up to the frequency appropriate for the location of the hair cell along the cochlear duct. As may be expected, for smaller diameter stereocilia with larger rootlet separations than were assumed in generating Fig. 5, the viscous forces on the stereocilia are decreased, and the corner frequencies correspondingly increased. Similarly, for larger diameter stereocilia with closer rootlet separations, the corner frequencies are decreased.

D. Dependence on material properties

The results discussed above are from a Young's modulus of both the stereocilia and the tip links of $E_{\text{cilia}} = E_{\text{tip}} = 10^7 \text{ N/m}^2$. Although this assumption is reasonable for biological materials (Wainwright *et al.*, 1976), these values are only estimates in the absence of experimental values. For purpose of comparison, these assumed material properties provide reasonable agreement with order of magnitude stiffness measurements reported in the literature (Table III, Szymko *et al.*, 1992, for review). In this model it can be seen that an increased material modulus of the tip links $E_{\text{tip}} > E_{\text{cilia}}$, causes both the corner frequency and the maximum frequency over which all of the stereocilia are in phase with the input force to increase. For a decreased material modulus $E_{\text{tip}} < E_{\text{cilia}}$, the range over which corner frequency solutions occur for varying bundle geometries is diminished. It is also interesting to note that viscous coupling provided between the stereocilia is sufficient such that amplitude and phase motion similar to that shown in Fig. 4 are provided when all linkages are absent, although the corner frequency is lowered and phase response begins earlier.



stereocilia, diameter, and rootlet separations which are similar to those of the guinea pig (Zetes, 1995). In addition, the frequency map reported in Greenwood (1990) is based on data from only the basal 60% of the cochlea, such that it may be inaccurate below 0.5 kHz.

In light of these difficulties, inspection of the available solutions in Fig. 5 indicates that the dimensions of the stereocilia bundle and corresponding filtering properties are consistent with the tonotopic arrangement of the cochlea. Motion of the stereocilia bundle relative to the apical surface of the hair cell exists without substantial attenuation at least up to frequencies corresponding to the location of the hair cell along the cochlear duct. The variation in stereocilia morphology along the cochlear duct may thus provide a collection of low-pass mechanoreceptors, arranged in order of increasing corner frequency across the auditory spectrum.

B. Analytical approximations and limitations

The equations derived above are for the linear motion of the stereocilia bundle. The range of validity of such a model is dependent on the assumptions of viscous dominated laminar flow and linear elastic stability of the bundle constituents. The hydrodynamic approximations discussed in the Appendix are valid for displacements of the tips of the stereocilia up to 1° (approximately 20 nm), at auditory frequencies of the chinchilla. Perhaps more limiting is the possibility of nonlinear contributions to the equations of motion from buckling of the tip links. Assad *et al.* (1991) and Assad and Corey (1992) demonstrate that there is a resting value of tension in the tip links. This resting tension provides a linear contribution of the tip links to the equations of motion for stereocilia deflections in both the tensile and compressive directions. Assuming that no other nonlinearity is apparent from ion channels resident in the stereocilia (Hudspeth, 1989), these equations of motion are valid for sinusoidal oscillation until the amplitude becomes sufficiently large such that the smallest net force during the cycle is compressive. This model is therefore reasonable for the linear response of the stereocilia bundle, at lower sound-pressure levels.

Solutions from this analysis have primarily addressed the issue of morphological variation as it affects stereocilia motion. For consistent comparison, the loading case and other parameters regarding the stiffness of the stereocilia and linkages have been held constant and expressed in their simplest form. In particular, the choice of an applied load at the tip of the tallest stereocilium is most appropriate for outer hair cell stereocilia since inner hair cell stereocilia are not attached to the tectorial membrane (Pickles, 1988, for review). The apparent rotational stiffness of the stereocilia has also been written in accordance with Euler–Bernoulli beam theory. To account for shearing deformation as observed in preparations for electron microscopy (Tilney *et al.*, 1983), similar equations can be written using Timoshenko theory (Peterson *et al.*, 1996; Zetes, 1995). When the material properties of the bundle constituents have been quantified experimentally, an analysis of such rigor coupled with the hydrodynamic solutions described above and alternate loading cases will provide a more accurate prediction of the stereocilia bundle frequency response. Further morphological stud-

ies will also allow a full-scale numerical study to provide understanding of the variation in stereocilia geometry along the cochlear duct in relation to Greenwood's frequency and hypotheses regarding mechanotransduction.

IV. SUMMARY

This paper introduces an analytical model to mathematically simulate the fluid–structure interaction of the stereocilia bundle for physiologically reasonable deflections and frequencies of excitation. Results from this model were used to investigate the frequency response of stereocilia bundles of the chinchilla cochlea, as quantified by tension in the tip links for the loading case of a force applied to the tip of the tallest stereocilium. It was found that the influence of the surrounding viscous endolymphatic fluid in proportion to the angular velocity of the stereocilia allows the amplitude of the tension in the tip links to remain constant for low frequencies of excitation. For higher excitation frequencies, viscous forces limit the motion of the stereocilia, thereby reducing the tension in the tip links in a fashion analogous to that of a low pass filter. Comparison of solutions for differing stereocilia geometries along the cochlear duct demonstrates that motion of the stereocilia exists without substantial attenuation at least up to frequencies appropriate for the location of the corresponding hair cell along the cochlear duct. This behavior is consistent with the cochlear tonotopic map. The variation in stereocilia morphology along the cochlear duct thus appears to provide a collection of low-pass mechanoreceptors, arranged in order of increasing corner frequency across the auditory spectrum. Further consideration with this model and similar analyses may provide insight into the functional implication of stereocilia arrangements and more specific comparison of hypotheses regarding mechanotransduction.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC00108 to Charles R. Steele.

APPENDIX

Consideration of the nondimensional form of the governing incompressible fluid equations demonstrates that the dominant hydrodynamic forces between the stereocilia are linear and viscous. Defining nondimensional independent variables for space x' , and time t' , and nondimensional dependent variables for speed u' , pressure p' , at a characteristic point in time, incompressible Navier–Stokes can be written

$$-E\nabla p' + \frac{1}{Re} \Delta \mathbf{u}' - \frac{1}{F} \nabla h' = S \frac{\partial \mathbf{u}'}{\partial t'} + \mathbf{u}' \cdot \nabla \mathbf{u}', \quad (\text{A1})$$

where E is the Euler number, Re is the Reynold's number, F is the Froude number, S is the Strouhal number, and the body force has been written as a gravitational force $-g\nabla h$, where h is the nondimensional elevation and g is the gravitational constant. These coefficients are given by

$$E = \frac{P_0}{\rho_0 U^2}, \quad \frac{1}{Re} = \frac{\mu}{\rho_0 U L}, \quad \frac{1}{F} = \frac{gL}{U^2}, \quad S = \frac{L}{TU}, \quad (A2)$$

where T , L , U , P_0 , and ρ_0 are characteristic values for the time, length, velocity, pressure, and density, respectively, and μ and g are physical values for viscosity and gravity.

For the case of the stereocilia bundle of the chinchilla, characteristic values are $L = 10^{-7}$ m, $T = 10^{-5}$ s, $U = 10^{-3}$ m/s, $P_0 = 10^{-3}$ N/m², $\rho_0 = 10^3$ kg/m³, $\mu = 10^{-3}$ kg/m s, and $g = 10$ kg m/s², where L is the separation between stereocilia shafts, T is the time for a stereocilium to reach its maximum velocity at the highest audible frequency (30 kHz, Greenwood, 1990), U is the maximum linear velocity of the tip of the stereocilia (approximately 10^{-6} m tall; Lim, 1980) for a deflection of 1° at the highest audible frequency, P is the pressure as estimated from Bernoulli's equation, and μ and ρ_0 are approximations of the material constants for the endolymphatic fluid from values for water.

Evaluating the coefficients gives $E = 1$, $1/Re = 10^4$, $1/F = 1$, and $S = 10$. It can thus be seen that the viscous contributions as quantified by the inverse of Reynold's number are three orders of magnitude larger than the inertia contributions as quantified by the Strouhal number at auditory frequencies. The nondimensional density and Froude number also demonstrate that the contribution of the nonlinear convection terms and gravitational body force are small. Hydrodynamic contributions for motion of the stereocilia within the endolymphatic fluid are therefore reasonably approximated by consideration of only linear viscous and pressure forces in the governing incompressible fluid equations.

- Allen, J. B. (1990). "Cochlear micromechanics—A physical model of transduction," *J. Acoust. Soc. Am.* **68**, 1660–1670.
- Assad, J. A., Hacohen, N., and Corey, D. P. (1992). "Voltage dependence of adaptation and active bundle movement in bullfrog saccular hair cells," *Proc. Natl. Acad. Sci. USA* **86**, 2918–2922.
- Assad, J. A., and Corey, D. P. (1992). "An active motor model for adaptation by vertebrate hair cells," *J. Neurosci.* **12**, 3291–3309.
- Assad, J. A., Shepherd, G. M. G., and Corey, D. P. (1991). "Tip-link integrity and mechanical transduction in vertebrate hair cells," *Neuron* **7**, 985–994.
- Authier, S., and Manley, G. A. (1995). "A model of frequency tuning the basilar papilla of the Toaky gecko, *Gekko gekko*," *Hearing Res.* **82**, 1–13.
- Corey, D. P., Hacohen, N., Huang, P. L., and Assad, J. A. (1989). "Hair cell stereocilia bend at their bases and touch at their tips," *Soc. Neurosci. Abstr.* **15**, 208a.
- Crawford, A. C., and Fettiplace, R. (1985). "The mechanical properties of ciliary bundles of turtle cochlear hair cells," *J. Physiol. (London)* **364**, 359–379.
- Duncan, R. K., Grant, J. W., and Peterson, E. H. (1993). "Finite element analysis of ciliary bundle stiffness," *Soc. Neurosci. Abstr.* **19**, 1579.
- Duncan, R. K., Hernandez, H. N., and Saunders, J. C. (1995). "Relative stereocilia motion of chick cochlear hair cells during high-frequency water-jet stimulation," *Aud. Neurosci.* **1**, 321–329.
- Flock, A., Flock, B., and Murray, E. (1977). "Studies on the sensory hairs of receptor cells in the inner ear," *Acta Oto-Laryngol.* **83**, 85–91.
- Freeman, D. M., and Weiss, T. F. (1990). "Superposition of hydrodynamic forces on a hair bundle," *Hearing Res.* **48**, 1–16.
- Geisler, C. D. (1993). "A model of stereociliary tip-link stretches," *Hearing Res.* **65**, 79–82.
- Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Hackney, C. M., and Furness, D. N. (1995). "Mechanotransduction in vertebrate hair cells: structure and function of the stereociliary bundle," *Am. J. Physiol.* **268**, C1–C13.
- Hudspeth, A. J. (1989). "How the ear's works work," *Nature (London)* **341**, 397–404.
- Hudspeth, A. J., and Corey, D. P. (1977). "Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli," *Proc. Natl. Acad. Sci. USA* **74**, 2407–2411.
- Jacobs, R. A., and Hudspeth, A. J. (1990). "Ultrastructural correlates of mechanoelectrical transduction in hair cells of the bullfrog's internal ear," *Cold Spring Harbor Symposia on Quantitative Biology*, Volume LV, pp. 547–561 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Lim, D. J. (1980). "Cochlear anatomy related to cochlear micromechanics. A review," *J. Acoust. Soc. Am.* **67**, 1686–1695.
- Ocvirk, F. W. (1952). "Short-bearing approximation for full journal bearings," NACA Tech. Note, 2808, Cornell University, NACA, Washington, October 1952, 61 pp.
- Pae, S. S., and Saunders, S. C. (1994). "Intra- and extracellular calcium modulates stereocilia stiffness on the chick cochlear hair cells," *Proc. Natl. Acad. Sci. USA* **91**, 1153–1157.
- Peterson, E. H., Cotton, J. R., and Grant, J. W. (1996). "Structural variation in ciliary bundles of the posterior semicircular canal," in *New Directions in Vestibular Research*, edited by S. M. Highstein, B. Cohen, and J. A. Büttner-Ennever (New York Academy of Sciences, New York), pp. 85–102.
- Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing* (Academic, San Diego, CA).
- Pickles, J. O. (1993). "A model for the mechanics of the stereociliar bundle on acousticolateral hair cells," *Hearing Res.* **68**, 159–172.
- Shotwell, S. L., Jacobs, R., and Hudspeth, A. J. (1981). "Directional sensitivity of individual vertebrate hair cells to controlled deflection of their hair bundles," *Ann. (N.Y.) Acad. Sci.* **374**, 1–10.
- Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (1993). "Electromechanical models of the outer hair cell," in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Nettern (World Scientific, Singapore), pp. 207–214.
- Strelhoff, D., and Flock, A. (1984). "Stiffness of sensory-cell hair bundles in the isolated guinea pig cochlea," *Hearing Res.* **15**, 19–28.
- Strelhoff, D., Flock, A., and Minser, K. E. (1985). "Role of inner and outer hair cells in mechanical frequency selectivity of the cochlea," *Hearing Res.* **18**, 169–175.
- Szymko, Y. M., Dimitri, P. S., and Saunders, J. C. (1992). "Stiffness of hair bundles in the chick cochlea," *Hearing Res.* **59**, 241–249.
- Tilney, L. G., Egelman, E. H., DeRosier, D. J., Saunders, J. C. (1983). "Actin filaments, stereocilia, and hair cells of the bird cochlea. II. Packing of actin filaments in the stereocilia and in the cuticular plate and what happens to the organization when the stereocilia are bent," *J. Cell Biol.* **96**, 822–834.
- Wainwright, S. A., Bigg, W. D., Currey, J. D., and Gosline, J. M. (1976). *Mechanical Design in Organisms* (Princeton U. P., Princeton, NJ).
- Zetes, D. E. (1995). "Mechanical and morphological study of the stereocilia bundle in the mammalian auditory system," Ph.D. thesis, Stanford University.
- Zetes, D. E., and Steele, C. R. (1997). "Mechanical models of the stereocilia bundle of the mammalian cochlea," in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poiner (World Scientific, Singapore).

Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers

Roger L. Miller,^{a)} John R. Schilling,^{b)} Kevin R. Franck,^{c)} and Eric D. Young
Center for Hearing Science and Department of Biomedical Engineering, The Johns Hopkins School
of Medicine, Baltimore, Maryland 21205

(Received 9 October 1996; revised 22 January 1997; accepted 23 January 1997)

A population study of cat auditory-nerve fibers was used to characterize the permanent deficits induced by exposure to 110–115 dB SPL, narrow-band noise. Fibers in the region of acoustic trauma (roughly 1–6 kHz) showed a loss of sensitivity at best frequency (BF) of about 50–60 dB and an increased tuning bandwidth. A correlation between weakened two-tone suppression and loss of sensitivity was found for fibers with BFs above 1 kHz. Single-fiber responses to the vowel /ε/ were recorded at intensities ranging from near threshold to a maximum of about 110 dB SPL. In normal cochleas, the temporal response patterns show a capture phenomenon, in which the first two formant frequencies dominate the responses at high sound levels among fibers with BFs near the formant frequencies. After acoustic trauma, fibers in the region of threshold shift synchronized to a broad range of the vowel's harmonics and thus did not show capture by the second formant at any sound level used. The broadband nature of this response is consistent with the broadened tuning observed in the damaged fibers, but may also reflect a weakening of compressive nonlinearities responsible for synchrony capture in the normal cochlea. © 1997 Acoustical Society of America. [S0001-4966(97)01106-5]

PACS numbers: 43.64.Sj, 43.64.Wn, 43.66.Ts [RDF]

INTRODUCTION

Previous studies of auditory-nerve pathophysiology following noise or ototoxic damage to the cochlea have demonstrated a loss of sensitivity at best frequency (BF) and a broadening of tuning curves (Kiang *et al.* 1976; Robertson and Johnstone, 1979; Liberman and Mulroy, 1982; Robertson, 1982). These changes have been related to specific damage to hair cells, especially to hair cell stereocilia (Liberman and Dodds, 1984). Diminished two-tone suppression has also been reported after damage to the organ of Corti (Schmiedt *et al.*, 1980, 1990; Salvi *et al.*, 1982). The effects of loss of sensitivity are straightforward, in that responses are not observed to weak stimuli, but the implications of broadened tuning and loss of two-tone suppression are more difficult to predict. In particular, the implications for responses to complex sounds, such as speech, have received little attention, although some studies have been reported (Geisler, 1989; Palmer and Moorjani, 1993).

Studies of responses to speech sounds in normal auditory-nerve fibers have demonstrated widespread synchrony to the formant frequencies (Young and Sachs, 1979; Sinex and Geisler, 1983; Delgutte and Kiang, 1984; Palmer *et al.*, 1986; Palmer, 1990). As sound level increases, the formants dominate the responses of fibers over a broad BF region and the responses of fibers with BFs near a formant

frequency are captured by the largest harmonic near the formant, meaning that they show phase locking only to that harmonic. It is this capture of responses by the formants that makes temporal representations of spectral shape so robust (Young and Sachs, 1979). Both the spread of synchrony and the capture phenomenon reflect nonlinear signal processing by the cochlea, in that they are not predictable from a fiber's tuning properties.

In this paper, we report the results of a population study of the response to the vowel /ε/ in auditory-nerve fibers of cats with noise-induced cochlear damage. These fibers show dramatic differences from those in normal cats in terms of their temporal response to the vowel. In particular, synchrony is more broadband, in that the response of a single fiber shows phase locking to many harmonics of the vowel. The nonlinear suppression implicit in the capture phenomenon is thus less evident after acoustic trauma. These results provide a neurophysiological basis for the diminished quality of speech perception demonstrated by human listeners with sensorineural hearing loss and make clear the difficulties that are encountered in designing prosthetic systems, such as hearing aids, to restore normal cochlear frequency analysis.

I. METHODS

A. Acoustic trauma

The acoustic trauma was induced by presenting a synthesized 50-Hz noise band, centered at 2 kHz, for 2 h at a level of 110 or 115 dB SPL in free field. The exposure was performed with the animal's head restrained and positioned directly beneath two loudspeakers. In preparation for the noise exposure, animals were anesthetized by an injection (im) of 2 mg of acepromazine maleate mixed with 200 mg of ketamine. Additional doses were administered as needed. At-

^{a)}Corresponding author address: Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, 505 Traylor Building, 720 Rutland Avenue, Baltimore, MD 21205. Electronic mail: rmiller@bme.jhu.edu

^{b)}Current address: Pritzker School of Medicine, University of Chicago, Chicago, IL 60637.

^{c)}Current address: Department of Speech and Hearing Sciences, University of Washington, Seattle, WA 98195.

ropine was also given (0.1 mg, im) to control mucus secretions. The animals were allowed to recover from the exposure for at least 44 days. The average recovery period was 57 days, with 80 days as the maximum recovery period. A previous study by Miller *et al.* (1963) used psychophysical testing of cats subsequent to a similar, 2-h acoustic exposure and showed nearly stable thresholds after 18.7 days and little change in threshold after one month.

Results are reported here for two groups of animals: (1) a control group of ten animals, consisting of seven unexposed animals and three cats that were exposed to the trauma at 100 dB SPL, but showed no deficits in their auditory-nerve responses; and (2) an impaired group of seven cats exposed as described above. The cats in the impaired group showed similar threshold shifts, so that they could be combined into a single pool. Twelve additional cats were exposed, but showed significantly different threshold shifts and so were excluded from the impaired pool. The responses in these additional animals were not qualitatively different from those in the impaired pool.

B. Electrophysiology

In preparation for recording from the auditory nerve, animals were anesthetized by injection of xylazine (0.2 mg im) followed by ketamine (100–200 mg im). An areflexic state of anesthesia was maintained for the duration of the experiment by intravenous injections of pentobarbital (≈ 12 mg/h IV). Physiological saline (≈ 1 ml/h IV) and lactated Ringer's (≈ 10 ml/24 h IV) were given to prevent dehydration. Atropine (0.1 mg im) was given every 24 h to control mucus secretions. A tracheotomy was performed to maintain an open, low-resistance airway, and to minimize respiratory acoustic artifacts. The bulla was vented with a 40 cm length of polyethylene tubing (PE-90) to prevent the buildup of negative pressure in the middle ear (Guinan and Peake, 1967). Rectal temperature was maintained near 38.5 °C with a feedback-controlled heating pad.

Recordings were made in an electrically shielded, double-walled soundproof room (IAC-1204 A). Micropipettes of 10–30 M Ω were filled with 3M KCl and placed in the auditory-nerve under visual control after retraction of the cerebellum medially. Single fibers were isolated by driving the electrode in 1–2 μ m steps while presenting a broadband noise search stimulus (about 35 dB *re*: 20 μ Pa/ $\sqrt{\text{Hz}}$). We continued to record from fibers in an animal until satisfactory recordings could no longer be obtained (usually 2–3 days). The state of the cochlea was monitored by plotting threshold versus BF and looking for sudden shifts in the thresholds of the most sensitive fibers. In normal animals, thresholds were stable (± 10 dB) over the 2–3 day recording period; in impaired animals, a long-term trend of increasing thresholds (up to 40 dB) was frequently observed (4/7 animals) in fibers with BFs < 1 kHz only. Despite the threshold shift, these data are included in the impaired pool.

C. Stimuli

The following data were taken from fibers in both impaired and control animals: (1) the frequency threshold curve

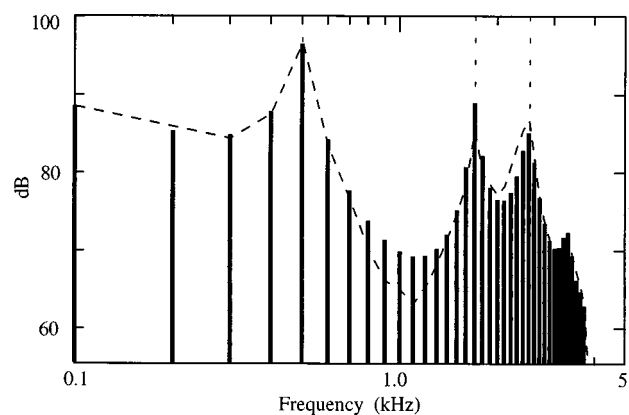


FIG. 1. Line spectrum for the steady-state vowel / ϵ / as synthesized by a cascade synthesizer, following Klatt (1980). The fundamental frequency (F_0) was 100 Hz and the formant frequencies and bandwidths used for synthesis were: F_1 (0.5/0.2 kHz), F_2 (1.7/0.2 kHz), F_3 (2.5/0.2 kHz), F_4 (3.3/0.25 kHz), and F_5 (3.7/0.2 kHz). The synthesizer output, as shown here, has been filtered by an approximation to human external ear acoustics (Wiener and Ross, 1946). Positions of the first three formants are indicated by vertical dashed lines. A representative spectral envelope of the stimulus, as presented at the tympanic membrane of one animal, is shown with a dashed line.

(FTC); (2) a 10 sample of spontaneous activity from which the spontaneous rate (SR) was determined; (3) responses to a BF tone 50 dB above threshold; (4) a measure of two-tone suppression; and (5) responses to the syllable “besh.” The FTC was generated by an automated procedure following Kiang *et al.* (1970) which tracked the sound level necessary to produce a one spike increase in response during a 50-ms tone burst. The two-tone suppression measure is described below (Fig. 5).

The syllable “besh” was presented 60–100 times at a particular sound level, with a one second interstimulus interval. Generally, more than 600 spikes were recorded in response to the steady-state vowel portion, although this was not possible for fibers with threshold shifts rendering them unresponsive to the stimulus. The stimulus was presented at 3 sound levels, 20 dB apart. The vowel’s power was computed as the sum of power in the first 50 harmonics of its fundamental (100–5000 Hz) and corrected for the average power output of the acoustic driver measured near the cat’s eardrum for each experiment. The minimum vowel level was near 30 dB SPL for control animals and near 70 dB SPL for the impaired animals.

The “besh” stimulus was synthesized with a Klatt synthesizer (Klatt, 1980), implemented in the ACSL simulation language (Franck, 1994). Although data were taken during the whole syllable (300 ms), only responses to the vocalic portion will be described here. The central 100 ms of the syllable consisted of a steady-state, periodic / ϵ /, whose spectrum is shown in Fig. 1. Formant frequencies (F_1 , F_2 , etc.) were chosen to be near the values used by Young and Sachs (1979), but were modified to be equal to a harmonic of the 100-Hz fundamental frequency (F_0). Formant frequencies and bandwidths are given in the caption of Fig. 1. The stimulus synthesis included filtering by the external ear transfer

function of the human head (Wiener and Ross, 1946) to make the vowel's spectral shape at the cat's tympanic membrane approximate that at the human tympanic membrane.

Sound stimuli were delivered via a closed acoustic system similar to that designed by Sokolich (1977). In our system, the electrostatic driver was replaced with 5-in. dynamic speaker in order to increase the maximum tone presentation level to about 125 dB SPL (Schilling, 1995). The frequency response for this system was relatively flat (± 6 dB) up to 10 kHz, beyond which its gain decreased dramatically. This upper limit was sufficient for the study of auditory-nerve fibers with BF's within the spectral range of the speech stimulus. The spectrum envelope of the vowel, after modification by the frequency response of the acoustic system in one experiment, is shown by the dashed line in Fig. 1.

D. Analysis

The neural response to speech stimuli was computed as the average poststimulus time histogram (PSTH), with a bin width of 20 μ s. Only responses to a segment within the steady-state vowel portion of the stimulus are reported here. The frequency response of the fibers was analyzed by applying an 81.92-ms Hamming window $w(n)$ to the PSTH $p(n)$, taking the Fourier transform, and computing synchronized rate as follows:

$$R(kf_t) = \left| \sum_{n=0}^{N-1} w(n)p(n)e^{-j2\pi kn/N} \right| / \sqrt{N \sum_{n=0}^{N-1} w(n)^2}. \quad (1)$$

The $R(kf_t)$ are the synchronized rates at frequencies kf_t , where f_t is the frequency resolution of the analysis (1/81.92 ms = 12.2 Hz). The units of $R(kf_t)$ are spikes/s, because the PSTH $p(n)$ was normalized to have units of spikes/s. The term in the denominator in Eq. (1) is a correction for the attenuation of the signal by the window function $w(n)$.

A second measure, the power ratio, was used to describe the degree to which synchrony to a particular formant dominates the response. $PR(F_x)$ is defined as the sum of the power in the response at the frequency of formant F_x and its harmonics, divided by the total power in the response. Because phase locking in cat auditory-nerve fibers is not observed above about 5 kHz (Johnson, 1980), the summations are limited to frequency components below 5 kHz:

$$PR(F_x) = \frac{\sum_{m=1}^u R^2(m \cdot F_x)}{\sum_{n=1}^v R^2(n \cdot F_0)},$$

with $u \leq 3$ and $(u \cdot F_x) \leq 5$ kHz,

and $v = 50$, so that $(v \cdot F_0) \leq 5$ kHz, (2)

where F_0 is the fundamental frequency of the vowel (100 Hz). Note that the summations are taken over harmonics of the stimulus fundamental frequency F_0 only; for a periodic stimulus, intermediate frequencies contribute only noise to the analysis. The summation in the numerator is limited to the formant frequency and its second and third harmonics, which are the dominant higher order rectifier distortion products in response to the vowel (Young and Sachs, 1979).

II. RESULTS

A. Effects of acoustic trauma on FTC threshold and tuning

Data taken from the control and impaired groups of animals will be compared throughout this paper. Figure 2 shows examples of FTCs for fibers in the two groups. Figure 2A shows FTCs for 11 fibers from one control animal. These display the properties generally seen in normal auditory-nerve fibers: low thresholds and narrow tip regions (Evans and Wilson, 1973; Liberman, 1978; Schmiedt *et al.*, 1990). Figure 2B shows 10 FTCs from an impaired animal. At low (< 0.8 kHz) and high (> 6 kHz) frequencies these curves approach normal values. However, in the vicinity of the exposure center frequency, 2 kHz, they show severely elevated thresholds and a loss of their tip regions, typical of acoustic trauma (Kiang *et al.*, 1976; Liberman and Kiang, 1978; Liberman and Mulroy, 1982; Robertson, 1982; Liberman and Dodds, 1984).

In order to obtain a dense sampling of fibers, the control and impaired data pools were generated by combining data across cats. This is a valid operation only if the hearing of each animal was comparable. Here we show that the thresholds were similar for animals in each group. A contour passing through the lowest FTC thresholds was defined in each animal (e.g., the dashed lines in Fig. 2A and B). Plots of these best-threshold contours are shown for all of the animals contributing to the control group in Fig. 2C and for all animals in the impaired group in Fig. 2D. The thresholds are similar, within 10 dB of each other in the control pool and within 20 dB for the impaired pool, suggesting that the groups are reasonably homogeneous. Other measures characterizing the effects of acoustic trauma, such as tuning bandwidth (Figs. 3B and 4B) and two-tone suppression (Fig. 6) were also compared and were similar among the animals making up the two groups (Schilling, 1995). Finally, the results from the analysis of the vowel responses (shown in Figs. 11–14) were similar among animals within each pool.

Figure 3A shows single-fiber sensitivity at BF for fibers in the control pool; along with a dashed line marked LBTC which shows Liberman's best-threshold curve, obtained for chamber-reared cats which had not been exposed to acoustic trauma (Liberman, 1978). The symbols show fibers in the different SR classes; the most sensitive fibers are in the high SR category, as expected.

Thresholds in our control animals are significantly higher than the LBTC at low frequencies and are slightly lower in the vicinity of 3 kHz. These differences result partially from differences in treatment of the middle ear; the bulla was closed and vented in our preparation, as opposed to open in Liberman's cats. With a closed bulla, the absolute sensitivity is decreased by about 5 dB at BFs below 1 kHz and there should be a middle-ear cavity resonance near 3 kHz (Guinan and Peake, 1967). All animals included in the impaired pool and half of the animals in the control pool were obtained from one breeder (Liberty Labs). In previous experiments with unexposed animals from this breeder, we have noticed low-frequency hearing losses like those shown in Fig. 3. Because of these differences between our experi-

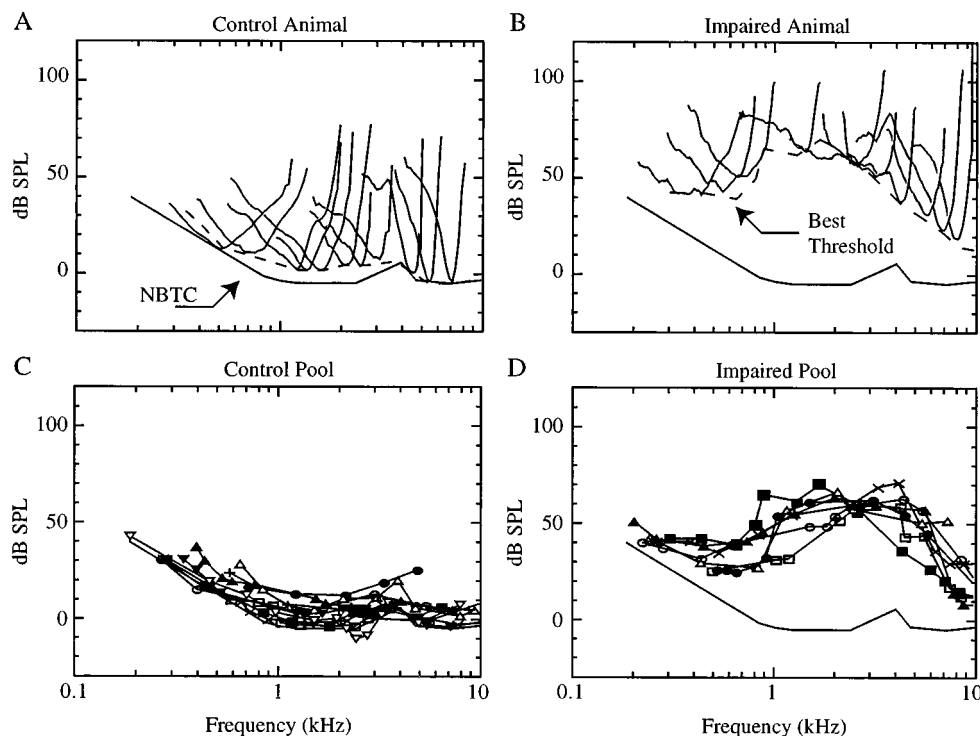


FIG. 2. Examples of FTCs from a normal, control animal (A) and from an impaired animal (B). For each animal, a best-threshold contour (dashed lines) was defined by connecting the tips of the FTCs from fibers with the lowest thresholds in each range of BFs. These curves were actually drawn from many more FTCs than are shown here. Best-threshold contours are compared for the ten animals in the control group (C) and for the seven animals in the impaired group (D). The line marked NBTC in A shows a smoothed version of the best threshold contour for all animals in panel C.

ments and those of Liberman, we defined a new best threshold curve, labeled NBTC in Figs. 2–4, to describe the best thresholds expected within our data; the NBTC was drawn as a sequence of straight line segments fitting the lowest thresholds in the control pool.

Figure 3B shows the Q_{10} measure of FTC width for the control pool. The dashed lines show the upper and lower boundaries of the values reported by Liberman (1978) for chamber-raised cats and the solid lines show the extent of Q_{10} values measured by Evans and Wilson (1973). Our values are in satisfactory agreement with these previous studies. Figure 3C shows a histogram of single-fiber BFs in the control pool; note the bias toward fibers with BFs near 2 kHz. Our electrode was generally placed at a point on the nerve where the BFs of the fibers isolated started near 0.5 kHz and then increased as the electrode was advanced. Tracks were terminated when fibers with BF below 10 kHz could not be found, and the electrode was replaced in a region where we expected to find 2-kHz fibers. This protocol would be expected to skew the distribution of isolated fibers towards bins near 2 kHz, as seen.

Figure 4 shows the thresholds, Q_{10} 's and BF distribution of the fibers included in the impaired pool. There is some uncertainty in the assignment of BFs to fibers following acoustic trauma. In this paper, we use the frequency of the local minimum threshold that is nearest to the high-frequency slope portion of the FTC, as opposed to the nominal BF. Liberman (1984) has shown that this convention gives the best agreement with the basilar membrane fre-

quency map following acoustic trauma. In Fig. 4A, single-fiber thresholds show a marked loss of sensitivity in comparison to the NBTC. Threshold shift was greatest for fibers in the 1–7 kHz range. Figure 4B shows that tuning is broadened in the region of acoustic trauma, especially over the BF range from 1 to 5 kHz. Most fibers with BFs greater than 5 kHz had Q_{10} values within the normal range indicating that less severe damage occurred at these higher frequencies. The distribution of fiber BFs, shown in Fig. 4C, did not show the same excess of fibers near 2 kHz as was seen in the control group (Fig. 3C), even though the same search strategy was used in the impaired cats.

Fibers are grouped into the usual low, medium and high SR groups in Figs. 3 and 4. Although fibers of all three SR groups are observed in impaired animals, the distribution of SR is different in impaired and normal animals. Our results (not shown) are the same as those reported previously (Fig. 19 of Liberman and Kiang, 1978). There is a decrease in the number of very high SR fibers ($>50/s$), an increase in the number of low and medium SR fibers and an overabundance of fibers with SR between 10 and 30/s.

B. Effects of acoustic trauma on two-tone rate suppression

The paradigm used to assess the extent of two-tone rate suppression is illustrated for three fibers from the control pool in Fig. 5. Two rate-versus-level functions were constructed simultaneously. A BF tone fixed at 10 dB above

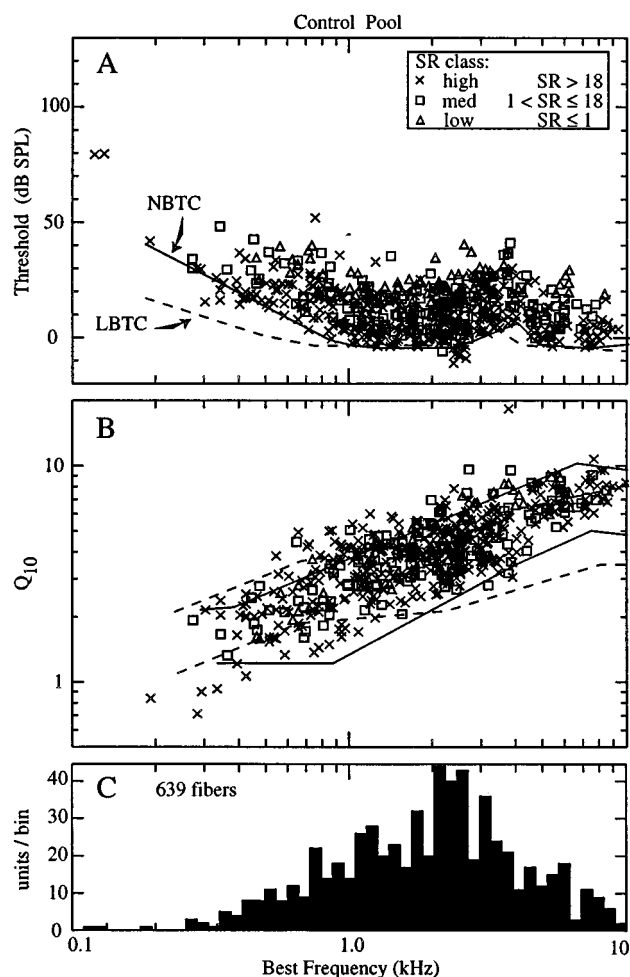


FIG. 3. Measures of tuning curve characteristics for fibers in the pool of normal hearing cats indicate normal sensitivity and tuning. Symbols denote SR class for the fiber, as shown in the legend. (A) Distribution of thresholds at BF. The solid line labeled NBTC describes the best thresholds for animals used in this study; it differs from the best thresholds in a previous study (LBTC, Liberman, 1978) at frequencies below 1 kHz and near the 3 kHz notch (see text). (B) Tuning curve width, measured as Q_{10} (BF divided by FTC bandwidth 10 dB *re*: threshold), was within the range of values obtained previously by others (dashed lines from Liberman, 1984; solid lines from Evans and Wilson, 1973). (C) Histogram of BFs for the fibers in this pool. The peak near 2 kHz indicates that our recording efforts were biased toward finding fibers near the second formant frequency.

threshold was alternately presented alone and in combination with a suppressor tone. The excitor or excitor-plus-suppressor were presented as 400-ms stimuli with an inter-stimulus period of 1 s. The level of the suppressor tone was changed in 2-dB steps over a maximum range of 100 dB. The upper stimulation limit was kept below ≈ 115 dB SPL and testing was generally stopped when the combination of BF and suppressor tone began to drive the fiber at a higher rate than the BF tone alone. Two suppressor tone frequencies were tested, one at $1.2 \times \text{BF}$ and the other at $0.25 \times \text{BF}$. Two-tone rate suppression is seen as a reduction in rate in the presence of the suppressor, for suppressor levels between 50 and 55 dB SPL in the case of Fig. 5A. Suppression can be clearly seen in Fig. 5B which shows the rate data of Fig. 5A replotted as fractional response, the ratio of the driven rate (total rate minus SR) to the excitor+suppressor combination divided by the driven rate to the excitor alone (Abbas and

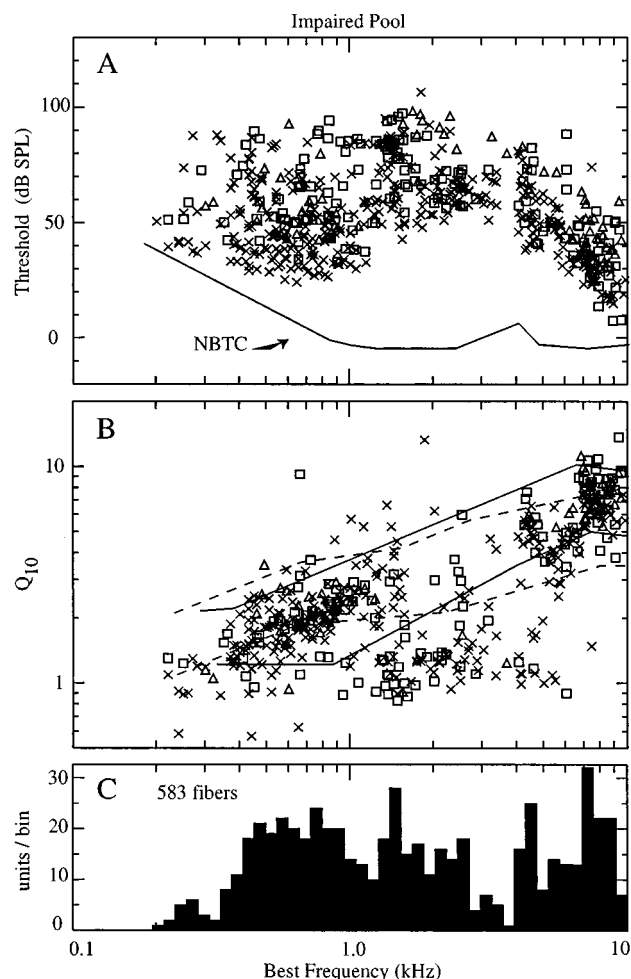


FIG. 4. FTC characteristics for fibers in the pool of impaired cats show elevated thresholds (A) and broadened tuning (B) in the 1.5–4 kHz BF range; these abnormalities are consistent with acoustic trauma which includes outer hair cell damage. Solid and dashed lines are the same as in Fig. 3. The BF distribution (C) shows that fewer fibers could be found in the region of the acoustic trauma, suggesting an inner hair cell loss.

Sachs, 1976). As the suppressor increased in level and the response to the excitor was suppressed, fractional response decreased to near zero. The rate and fractional response increased again as the fiber began responding to the suppressor tone. To quantify two-tone rate suppression, the minimum of the fractional response was estimated by fitting the fractional response trend with a second-order polynomial and using the minimum of that polynomial. This measure, *minimum fractional response* (MFR), is zero for complete rate suppression and one for no rate suppression. Data from two additional control fibers are included in panels C–F to illustrate variability in the fractional response contour. The polynomial fit provided both smoothing of the data and an objective estimation of the minima.

Figure 6 shows results for MFR in the normal and impaired pools. To show the relationship between two-tone suppression and the extent of hearing impairment, MFR is plotted against a measure of threshold shift. Because threshold varies with BF and SR as well as with hearing loss, a

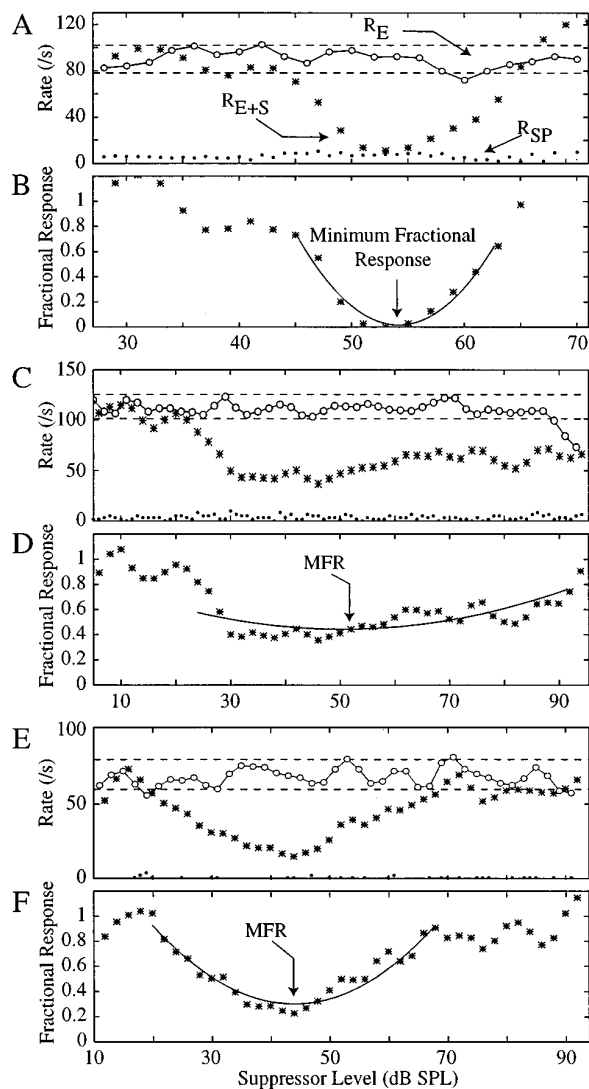


FIG. 5. Data from three fibers in the control pool illustrate the variability in the measure of two-tone suppression strength across the population of fibers. (A) Rate vs level functions were used to measure the strength of two-tone rate suppression. The rate functions show rate when the fixed-level excitator tone was presented alone (unfilled circles, R_E) and when the suppressor tone was present with the excitator (asterisk, R_{E+S}). The excitator tone was at BF (4.51 kHz) and was fixed at 10 dB above threshold; the suppressor tone was at $0.25 \times \text{BF}$ (1.13 kHz) and its level was increased in 2-dB steps. R_{E+S} is plotted vs suppressor tone level; R_E shows rate to the fixed-level excitator alone during control trials interleaved with the excitator+suppressor trials. Dashed lines show 1.5 standard deviation confidence limits for the rate to the excitator alone. The dotted line indicates spontaneous rate (R_{SP}) during the 0.6-s interstimulus interval for this medium SR fiber. (B) Fractional response, defined as $(R_{E+S} - R_{SP}) / (R_E - R_{SP})$, plotted versus suppressor level for the same fiber. Fractional response is one for no suppression and zero for complete suppression. The solid line shows a second-order polynomial fit to the data over the range where R_{E+S} is below the 1.5 standard deviation limit in (A). Minimum fractional response (MFR) is the minimum of this polynomial. (C), (D) Similar data for a medium SR fiber (BF=2.07 kHz) tested with a suppressor tone above BF (2.48 kHz) (E), (F). Similar data for a low SR fiber (BF = 1.93) with a 2.32-kHz suppressor tone.

relative measure of threshold shift was used. For each experiment, best-threshold contours were estimated as shown in Fig. 2A and B. These contours provide a measure of the threshold shift of the most sensitive fibers at each point on the cochlea; this is a direct measure of the degree of damage

at a point on the cochlea and is independent of the normal variation in threshold observed across SR groups. To remove the variation in threshold with BF, the best-thresholds were expressed in terms of shift from the NBTC, giving the *best threshold difference* (BTD). The abscissae in Fig. 6 show the BTD at the BF of each fiber.

The data in Fig. 6 are for fibers with BFs above 1 kHz only. This limited range of BFs was used for two reasons. First, the MFR measure of two-tone suppression ranges from zero to one for normal fibers with BFs above 1 kHz, but does not cover this full range in low BF fibers (other data on the same point are reviewed by Prijs, 1989). This BF dependence confounds the use of MFR as a measure of two-tone suppression here, since there is also a correlation between threshold shift and BF in the data. Second, the largest effects should be seen in fibers with BFs near the exposure frequency, i.e., in the 1–10 kHz range.

Figure 6A and C shows MFR data from the control pool for suppressor tones at $1.2 \times \text{BF}$ and at $0.25 \times \text{BF}$, respectively. The BTDs for these fibers are small, as is expected from Fig. 2. MFR varies over the full range from zero (maximal suppression) to one (no suppression). Note that all but one of the fibers in the low and medium SR groups (triangles and squares, respectively) show some degree of two-tone suppression, whereas many fibers with high SR show no suppression (MFR=1). To quantify this point for the suppressor tone at $1.2 \times \text{BF}$, 61% of low and medium SR fibers have $\text{MFR} < 0.6$ (which is the approximate median MFR), whereas only 19% of high SR fibers do; the difference is significant ($p < 0.002$, χ^2 test). When the suppressor tone was presented at $0.25 \times \text{BF}$, 90% of low and medium SR fibers have $\text{MFR} < 0.6$ (the approximate median MFR), whereas only 38% of high SR fibers do; this difference is also significant ($p < 0.001$, χ^2 test). This result is consistent with previous studies (Palmer and Evans, 1982; Costalupes *et al.*, 1984; Cai and Geisler, 1996). Figure 6B and D shows data from the impaired pool. The BTD varies from 15 to 70 dB and there is a trend toward weakening of suppression as the BTD increases. The lines in Fig. 6B and D are regression fits to the impaired-fiber data; in both cases, there is a significant correlation of MFR with BTD ($p < 0.001$). Rate suppression does not, however, disappear with any degree of threshold shift up to maximum included in the impaired pool. The lines in Fig. 6A and C are regression fits to data from the control and impaired pools combined together. These lines have similar slopes (0.0058 and 0.0048/dB, respectively), which is an estimate of the dependence of two-tone suppression, as measured by MFR, on cochlear damage, as measured by BTD.

C. Changes in phase locking to a BF tone

The effect of acoustic trauma on synchrony to a BF tone was estimated by calculating the synchronization coefficient $S_{50 \text{ dB}}$, equal to the synchronized rate to a BF tone 50 dB above threshold divided by the average discharge rate. The onset response was eliminated by omitting the initial 10 ms of the response to the 200-ms tone. The results are illustrated for the normal (Fig. 7A) and the impaired (Fig. 7B) pools.

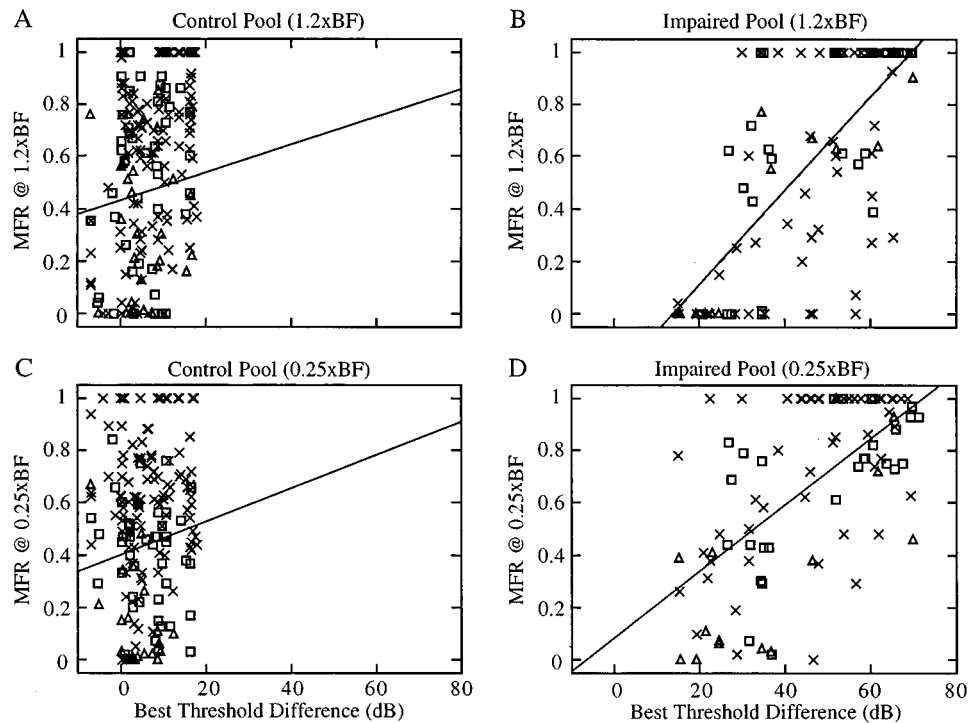


FIG. 6. The MFR measure of two-tone rate suppression, defined in Fig. 5, plotted against a measure of threshold shift (defined in text) for fibers in the control (A, C) and impaired (B, D) pools. Data are plotted for fibers with BFs above 1 kHz only. Panels A and B show results for the suppressor tones at $1.2 \times \text{BF}$; panels C and D show results for suppressors at $0.25 \times \text{BF}$. Symbols identify the SR group for each fiber, as defined in the legend of Fig. 3. Lines in panels B and D are regression lines for the impaired data: B: $\text{MFR} = -0.24 + 0.018 \times \text{BTD}$, $R = 0.71$, sign. at $P \ll 0.001$; D: $\text{MFR} = 0.09 + 0.013 \times \text{BTD}$, $R = 0.62$, sign. at $P \ll 0.001$. Lines in panels A and C show regression lines for the impaired and control pools combined: A: $\text{MFR} = 0.47 + 0.0048 \times \text{BTD}$, $R = 0.29$; C: $\text{MFR} = 0.45 + 0.0058 \times \text{BTD}$, $R = 0.41$.

Because the BF tone was limited to a maximum level of 115 dB SPL, impaired fibers with large threshold shifts could often not be tested at 50 dB above threshold (88/259 impaired fibers). Data from these 88 fibers are not plotted in Fig. 7B; however, phase locking in this subgroup of fibers shows the same trends as data that are plotted. The solid lines are drawn along the lowest $S_{50 \text{ dB}}$ values in the control pool (Fig. 7A); notice that there are few fibers below this line in the impaired pool (Fig. 7B). Thus the ability of fibers to phase lock is not impaired by the acoustic trauma; in fact, a few impaired fibers show synchrony that is higher than that seen in the control animals. For comparison, the dashed lines show the range of synchronization values obtained for normal fibers by Johnson (1980). These data differ from our values in that they show the maximum synchrony to a tone over a range of presentation levels. Because our data were obtained at a fixed level above threshold, they do not necessarily represent the maximum synchrony.

D. Changes in phase locking to the second and third formant frequencies

In the following sections we describe abnormalities in the phase locking of fibers grouped by their proximity to the major spectral features of the vowel. The frequency of the acoustic trauma, 2 kHz, was between the second (F_2) and third (F_3) formant frequencies. Therefore, the most extreme changes in the responses to the vowel were expected for

fibers with BFs near these formants. Figure 8 shows the FTCs for two normal fibers and two impaired fibers; the fibers in Fig. 8A have BFs near F_2 and the fibers in Fig. 8B have BFs near F_3 . Phase locking data from these four fibers are shown as examples in the next two figures. The FTCs of the impaired fibers in Fig. 8 show substantial threshold elevation; the FTC in Fig. 8B also shows a severe degradation in tuning. The method of estimating impaired fibers' BFs is illustrated by these examples. The arrows point to the BFs chosen for these fibers, which are at the nominal BF for the normal fibers, at the local minimum nearest the upper edge of the tuning curve for impaired fibers like the one in Fig. 8A, and at the point of departure of the FTC from a straight line fit to the high-frequency limb of the FTC for impaired fibers like the one in Fig. 8B (Liberman, 1984).

Figure 9 compares phase locking to the vowel for two fibers with BFs near F_2 (FTCs in Fig. 8A). The plots show the synchronized rate, $R(kf_i)$ in Eq. (1), in response to the vowel at three sound levels. The responses of the normal fiber are typical of those previously described (Young and Sachs, 1979). At low levels such as 31 dB SPL (Fig. 9C), fibers phase lock to harmonics of the vowel centered on its BF. Here the largest response is at F_2 , but synchrony to adjacent harmonics is also evident from the smaller peaks adjacent to the F_2 response. The Fourier transform also contains components near the second harmonic of F_2 , which are rectifier distortion products. As sound level increases (Fig. 9B and A), the response becomes more narrow band and is

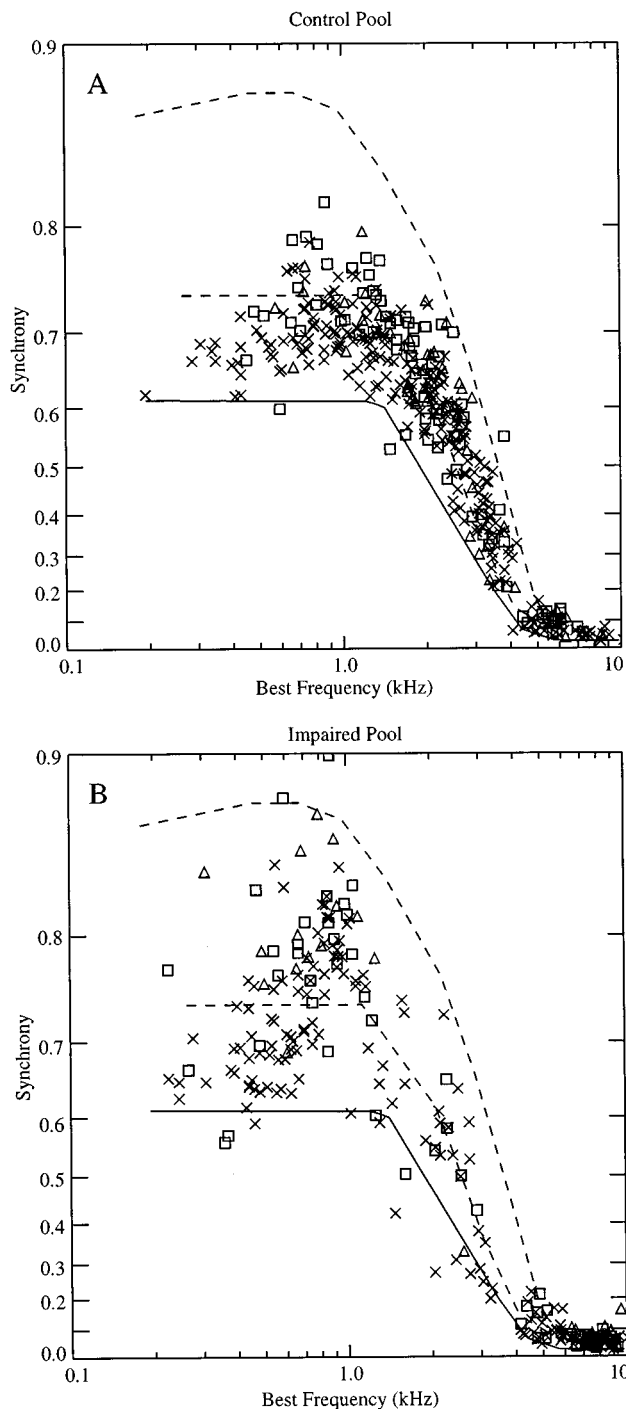


FIG. 7. Synchronization coefficient $S_{50\text{dB}}$ for a BF tone presented 50 dB above threshold, plotted against fiber BF. Points show data from all fibers in the control (A) and impaired (B) pools, except as noted in the text. The impaired pool (B) does not show a dramatic loss of synchrony, as compared to the control pool (A). The solid line shows the lower boundary of $S_{50\text{dB}}$ values from the control pool. Dashed lines show the range of maximum synchronization coefficient values for normal fibers (Johnson, 1980). The ordinate axis is an inverted logarithmic scale of $(1 - S_{50\text{dB}})$, along which synchrony has approximately constant variance (Johnson, 1980).

dominated by F_2 (1.7 kHz); phase locking to harmonics near F_2 disappears. We refer to this narrow-band response as capture of the fiber's response by F_2 .

The lower three panels (D–F) show synchrony for the impaired fiber, but at higher presentation levels, as required

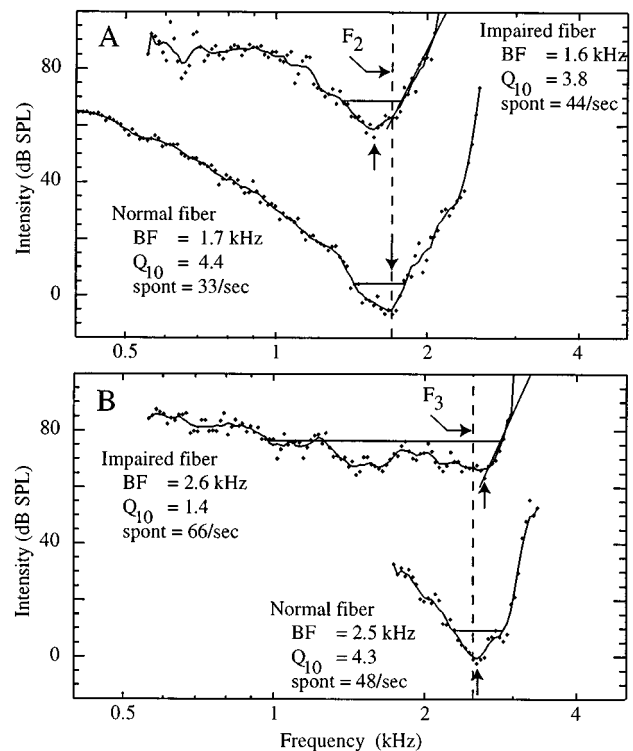


FIG. 8. FTCs for normal and impaired fibers with BFs near F_2 (A) and F_3 (B); each panel shows one normal and one impaired fiber. These are the fibers whose phase locking data are shown in Figs. 9 and 10. Horizontal lines indicate tuning curve width used to compute Q_{10} . Vertical dashed lines indicate the frequency of the nearest formant. The diagonal lines on the upper edge of the impaired FTCs show the high-frequency slopes; for impaired fibers with broad tuning, like the example in (B), BFs were chosen near the point of departure of the FTC from these straight lines. Arrows show the fibers' BFs. The normal fibers show thresholds close to Liberman's best thresholds and Q_{10} values within the normal range. The impaired fibers show a loss of sensitivity of about 60 dB; the Q_{10} of the impaired fiber in (A) is within the normal range, whereas the one in (B) is significantly decreased.

by the nearly 60-dB loss of threshold (Fig. 8A). At the lowest presentation level, 73 dB SPL (Fig. 9F), significant synchrony is present at a large number of harmonics, including the first formant. Synchrony to F_2 is only slightly stronger than to other harmonics. At higher presentation levels, synchrony to F_2 increases somewhat, but does not become the dominant feature, as occurred in the normal fiber. In fact, in this fiber the strongest synchrony at the highest level (113 dB SPL, Fig. 9D) is to the 16th harmonic of the vowel, approximately the fiber's BF.

At all three levels in Fig. 9D–F, there are two prominent response components, one at F_1 and the other at either F_2 or the 16th harmonic. If the fiber were responding to only these two stimulus components, then rectifier distortion products produced in the hair cell/auditory-nerve synapse would be expected (Young and Sachs, 1979). The largest distortion components are expected to be the harmonics of the primaries: $2 \times F_1$, $3 \times F_1$, $2 \times F_2$ and the first difference frequency, $F_2 - F_1$. Thus some of the broadband character of the responses shown in Fig. 9D–F can be explained simply as the distortion expected in a response to two primary com-

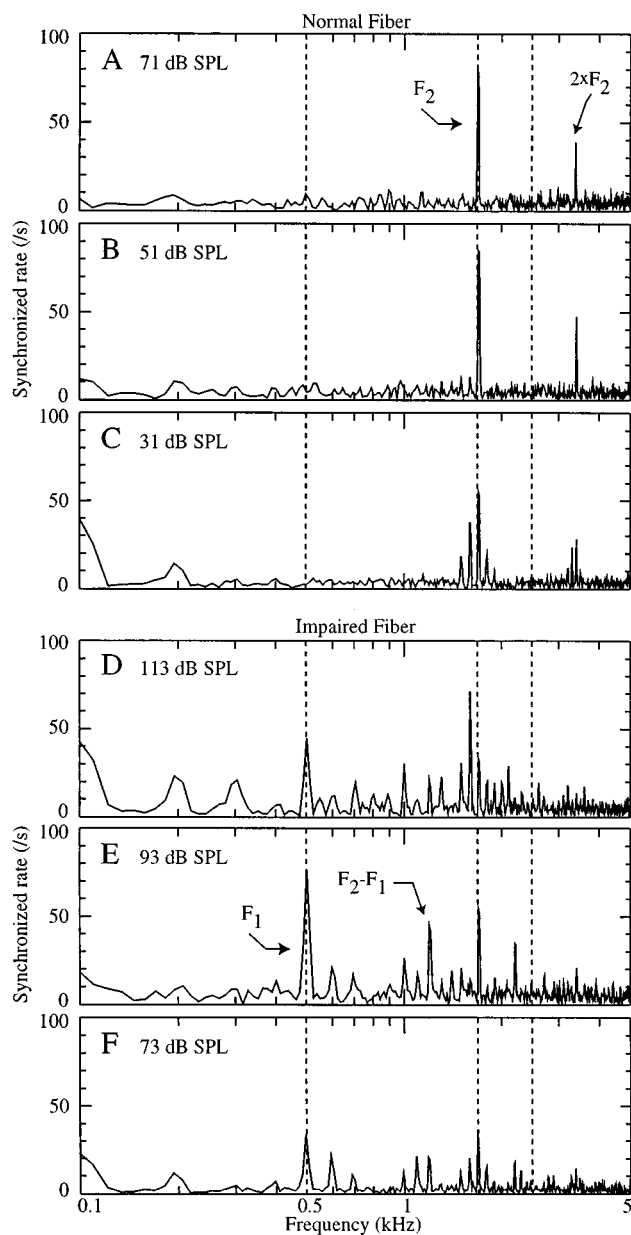


FIG. 9. Synchronized rates in response to the vowel for the two fibers with BFs near F_2 (FTCs in Fig. 8A). Plots are the magnitudes of the Fourier transforms of the PST histograms of responses to the vowel, $R(kf_i)$ in Eq. (1). The three plots for each fiber show the change in response pattern with increasing level; sound levels are indicated on the plots. Vertical dashed lines show the first three formant frequencies of the vowel. The normal fiber (A, B, C) shows synchrony capture by F_2 at the highest two levels. The impaired fiber (D, E, F) shows synchrony to a broad range of harmonics at all presentation levels, with no synchrony capture by F_2 . In fact, at the highest presentation level (D) synchrony to 1.6 kHz, the fiber's BF, is larger than synchrony to F_2 . Synchrony to F_1 is greater for the impaired fiber at all levels, as compared to the normal fiber.

ponents at F_1 and F_2 . However, detailed examination of the data in Fig. 9E and F shows that many prominent harmonics in the response (particularly harmonics 6, 7, 11, and 16) cannot be explained as distortion products from F_1 and F_2 (harmonics 5 and 17). Similarly, components at harmonics 7, 12, 13, and 17 cannot be explained as distortion products of the prominent primaries (harmonics 5 and 16) in Fig. 9D. Thus, the broadband character of the responses in the im-

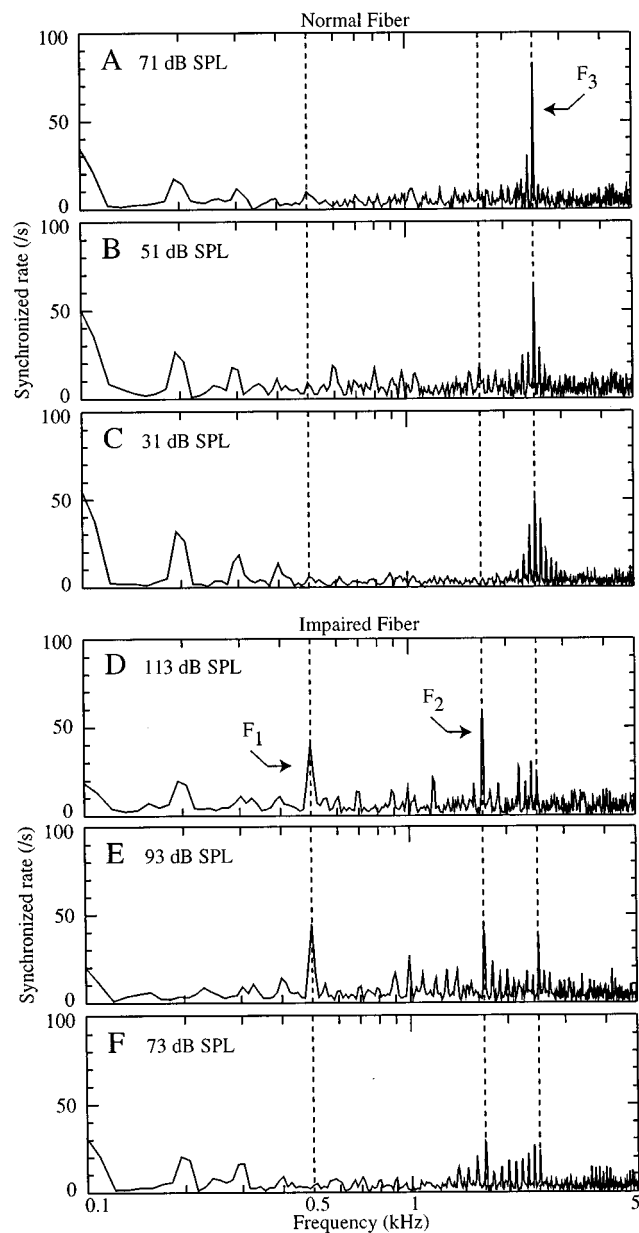


FIG. 10. Response synchrony for fibers with BFs near F_3 (FTCs in Fig. 8B) are shown in the same format as Fig. 9. In the normal fiber (A, B, C), synchrony rate to F_3 increases with presentation level but does not reach the level of synchrony capture. Synchrony patterns for the impaired fiber (D, E, F) show a broadband response; the response changes with presentation level and shows a poor representation of F_3 . Synchrony to F_2 , not F_3 , increases with level, and synchrony to F_1 is strong at the higher levels.

paired fiber reflects more than addition of a strong F_1 response to the expected F_2 response of the fiber.

Figure 10 shows data for the two fibers with BFs near 2.5 kHz, the third formant frequency; the organization of the figure is the same as Fig. 9. FTCs for these fibers are shown in Fig. 8B. In the normal fiber (Fig. 10A, B, and C), responses to F_3 and nearby harmonics were observed at all levels; although the synchronized rate to F_3 increased with presentation level, small synchrony to nearby harmonics remained. In contrast, phase locking to F_3 was not the dominant component of the response at any presentation level (Fig. 10D, E, and F). At the lowest presentation level, syn-

chrony to a broad range of harmonics in the $F2$ and $F3$ regions was observed. Increasing the presentation level by 20 dB (Fig. 10E) brought an increase in the synchrony to $F2$ and $F3$ and emergence of significant synchrony to $F1$. At the highest presentation level (Fig. 10D), synchrony to $F1$ and $F2$ remained strong and synchrony to $F3$ was no stronger than that to nearby nonformant harmonics. The most important result, however, is that synchrony to $F1$ and $F2$ remains greater than that to $F3$. This is in contrast to the normal case where synchrony to $F1$ and $F2$ did not rise above the noise level.

E. Power ratio measure of synchrony capture by formants

The examples shown in Figs. 9 and 10 are typical of the populations from which they are drawn. Figures 11–13 show the fraction of power in a fiber's response that is carried by a particular formant in terms of power ratio [Eq. (2)], plotted versus BF for the whole population of fibers. In general, the results from the control pool (panels A–C) show strong responses to the formants confined to the vicinity of the cochlear place appropriate to the formant frequency. By contrast, the impaired animals (panels D–F) show weaker responses to $F2$ and $F3$ at their place and a spread of the response to $F1$ and $F2$ along the cochlea.

Figure 11 shows the power ratio for $F2$, $PR(F2)$. As the average presentation level was increased from 28 dB SPL (Fig. 11C) to 49 dB SPL (Fig. 11B), synchrony to $F2$ increased for normal fibers with BFs equal to $F2$. At 69 dB SPL (Fig. 11A), the growth in synchrony for these fibers has saturated, but fibers at adjacent frequencies now show synchrony to this formant, as demonstrated by the increase in breadth of the response peak. Although there is no threshold value of the power ratio index which denotes synchrony capture, a value above 0.5–0.6 indicates strong synchrony to a formant. Impaired fibers (Fig. 11D–F) show a much lower $PR(F2)$ at all presentation levels. Note that the presentation levels in the impaired animals were higher than those in the normals, in order to compensate for the loss of sensitivity. This increase in level did not restore normal phase locking to $F2$. Instead, $PR(F2)$ remained below that of the normal fibers at the $F2$ place. The impaired fibers also show an increased tendency to synchronize to $F2$ over a wide range of BFs above $F2$, including the $F3$ place. Normal fibers show synchrony to $F3$, not $F2$, in this region, as will be shown in Fig. 13.

Power ratio for synchrony to $F1$ is shown in Fig. 12. There was relatively good agreement between the maximal $PR(F1)$ for normal and impaired fibers in the vicinity of the $F1$ place, including synchrony capture for fibers with BFs near $F1$ at the lowest two levels. There are two notable differences between the populations, however. First, synchrony to $F1$ decreases at the $F1$ place at the highest level (112 dB) in the impaired animals. Comparable levels were not presented in normal animals. This diminished $PR(F1)$ results from a loss of synchrony capture and is observed in normal animals at sufficiently high levels (Young *et al.*, 1996). The second difference is that impaired fibers in the

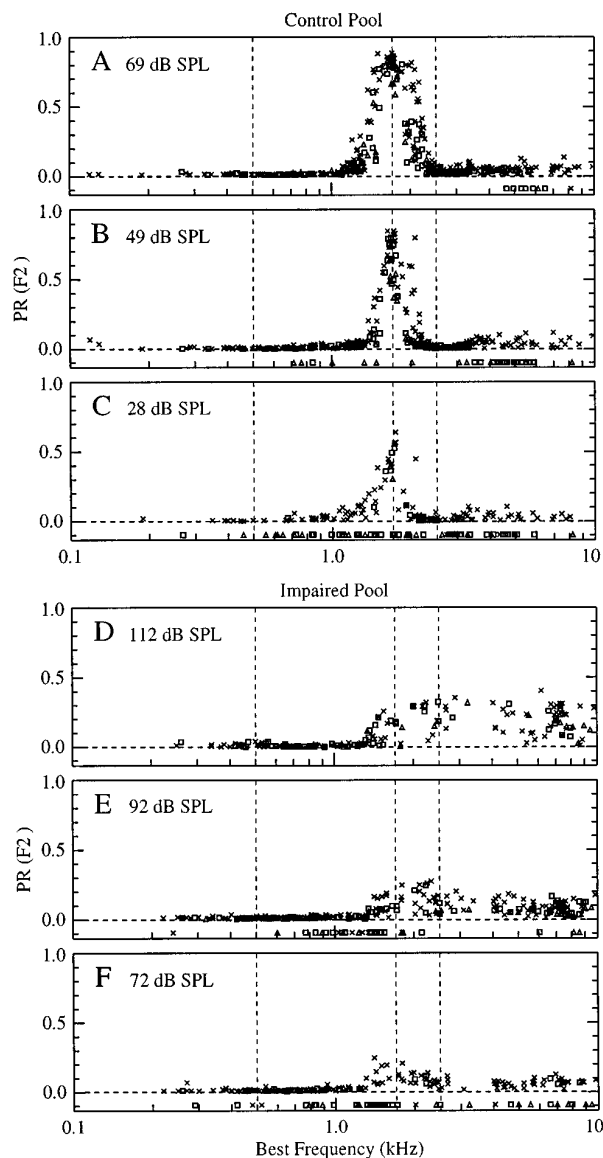


FIG. 11. Fraction of the power in fibers' responses to $F2$, $PR(F2)$ in Eq. (2), plotted against fiber BF. Data from the control pool (A–C) show large responses to $F2$ at the $F2$ place which do not spread far from that place. The response at the $F2$ place grows with stimulus level. Impaired fibers (D–F) showed diminished synchrony to $F2$ at the $F2$ place and show a broad spread of synchrony to higher BFs. The power ratio values do not approach the level which indicates synchrony capture, i.e., 0.5–0.6. For comparison, $PR(F2) = 0.86$ for both of the normal responses in Fig. 9A and B but is below 0.20 for the impaired responses in Fig. 9D and E. SR class is denoted by the symbol, as defined in the Fig. 3 legend. Vertical dashed lines show the frequencies of the first three formants. Symbols plotted along the abscissa show fibers which fired fewer than 200 spikes over 50 presentations, i.e., the stimulus was below that fiber's threshold. The sound pressure levels given in the legends are the average levels for the experiments making up each pool. Levels varied over a 14.9-dB range in the control pool and a 4.7-dB range in the impaired pool. Three animals in the control pool were not tested at the lowest presentation level, hence the average presentation level in panel C is not exactly 20 dB below that in panel B.

$F2$ – $F3$ region (Fig. 12D–F) showed an increase in $PR(F1)$. This change in response was also evident in Figs. 9 and 10.

The most dramatic reduction in formant representation subsequent to acoustic trauma is illustrated in Fig. 13. In the

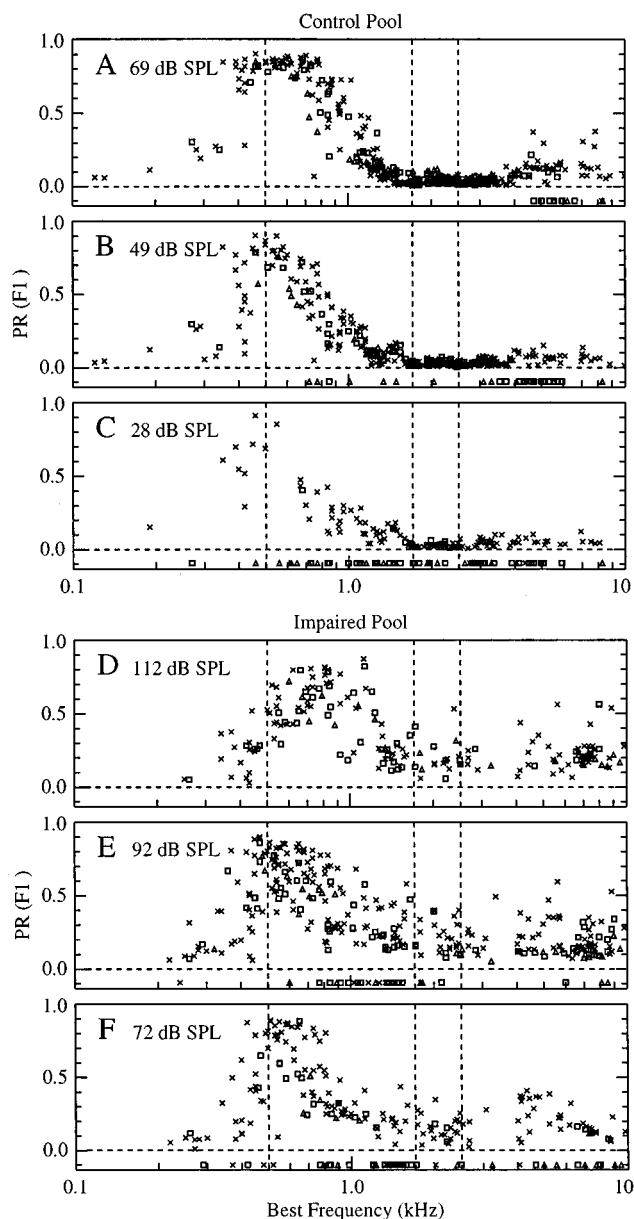


FIG. 12. Power ratio for F_1 plotted versus BF for normal (A, B, C) and impaired (D, E, F) fibers in the same format as Fig. 11. Normal fibers show a growth of synchrony to F_1 near the F_1 place, but little response to F_1 near the F_2 and F_3 places. Impaired fibers show substantial spread of response to F_1 to higher BF's at all levels.

normal fibers, Fig. 13A–C, clear synchrony to F_3 exists near the F_3 place; synchrony to F_3 increased with presentation level, although it usually did not reach the level of synchrony capture. In the impaired fibers (Fig. 13D–F), synchrony to F_3 remained at baseline levels for fibers of all BF. The lack of response to F_3 was not a simple result of the threshold shift of fibers near the 2.5-kHz place. Instead, fibers at the F_3 place were responding to F_1 and F_2 , as can be seen in Figs. 11 and 12.

The results above demonstrate phase locking to the harmonics at the formant frequencies. As shown with the single fiber examples in Figs. 9 and 10, the major effect of impairment on phase locking is an increase in the bandwidth of the response and a loss of synchrony capture by the formants.

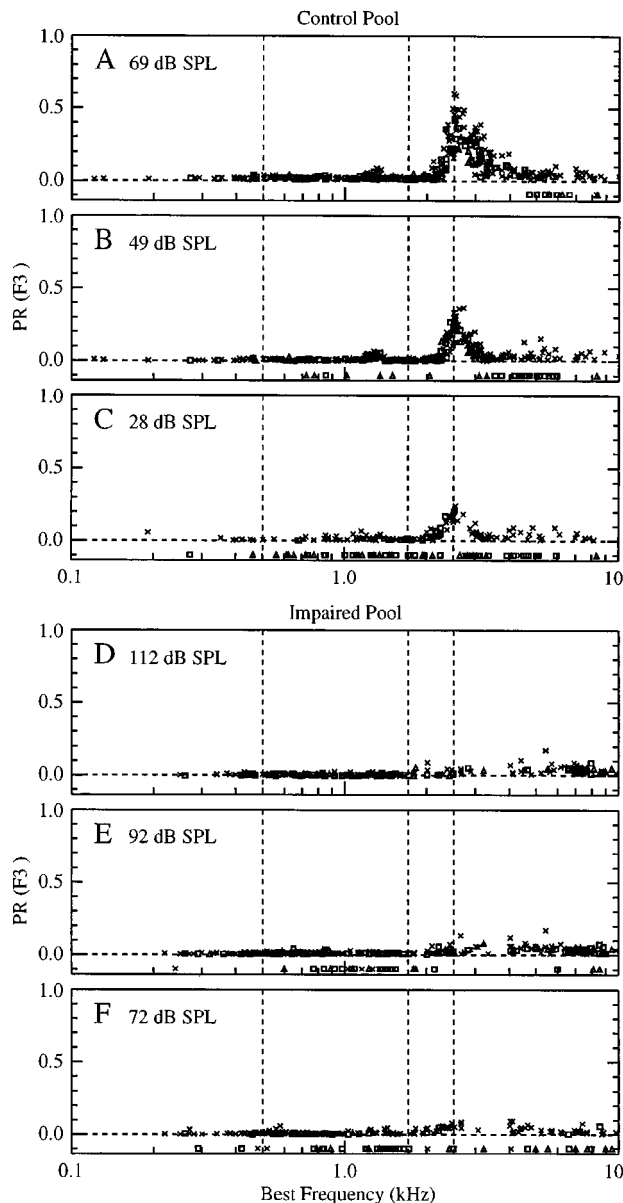


FIG. 13. Power ratio for F_3 plotted versus BF for normal (A, B, C) and impaired (D, E, F) fibers in the same format as Fig. 11. Normal fibers show a small F_3 synchrony peak near the F_3 place which grows with stimulus level, but it is not as large as the synchrony to F_2 at the F_2 place (Fig. 11). Impaired fibers show essentially no synchrony to F_3 . For comparison, the $PR(F_3)$ were 0.58 and 0.32 for the normal responses shown in Fig. 10A and B, respectively, and below 0.15 for the impaired responses shown in Fig. 10D and E.

The broadening of the bandwidth can be demonstrated by looking at the responses to frequency components in the trough between F_1 and F_2 . In normal animals, these are suppressed by the formant frequencies at higher stimulus levels; this suppression occurs to a lesser extent in the impaired animals. Figure 14 shows distributions by BF of the power ratio for stimulus components in the trough, computed as

$$PR(\text{trough}) = \frac{\sum_{m=6,7,8,9} R^2(m \cdot F_0)}{\sum_{n=11,13,14,16} R^2(n \cdot F_0)} \quad (3)$$

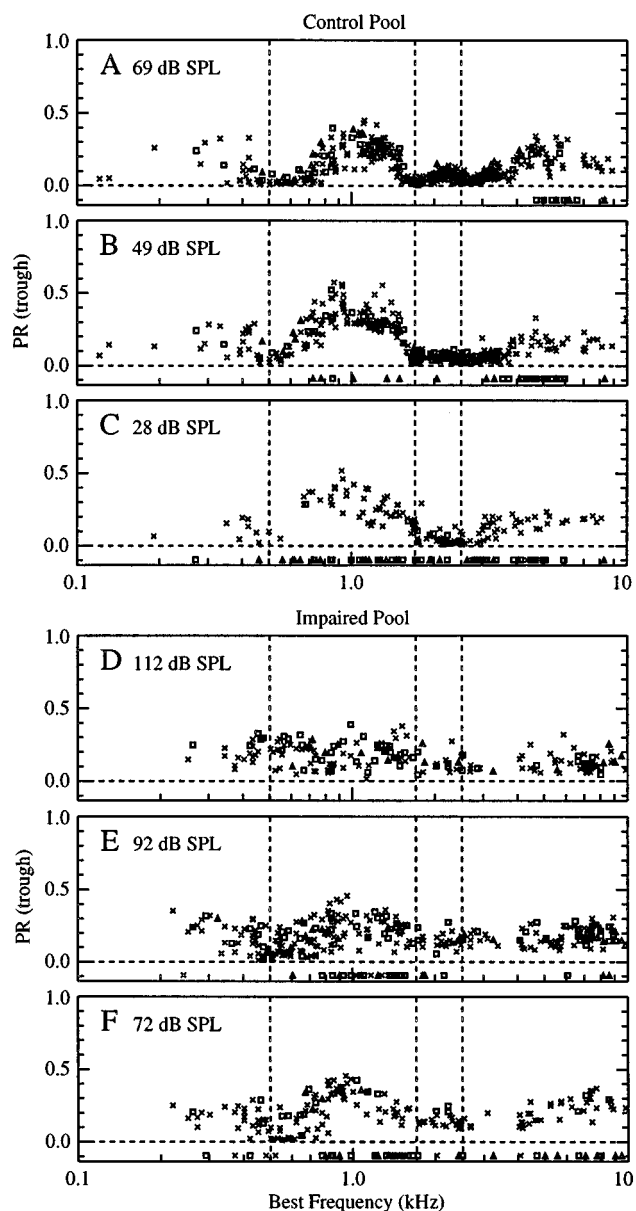


FIG. 14. Power ratio for trough frequencies, Eq. (3), plotted versus BF for normal (A, B, C) and impaired (D, E, F) fibers in the same format as Fig. 11. Normal fibers at formant places, indicated by vertical dashed lines, show little synchrony to these harmonics. Impaired fibers have a broadband response and thus respond to nonformant harmonics.

That is, $PR(\text{trough})$ is the fraction of the power in the response carried by all stimulus harmonics between $F1$ and $F2$ except those that are likely to be prominent rectifier distortion products from $F1$ and $F2$ (i.e., $2 \times F1 = 10\text{th}$, $3 \times F1 = 15\text{th}$, and $F2 - F1 = 12\text{th}$ harmonics). In normal fibers (Fig. 14A–C), some response to this range of frequencies occurs in fibers with BFs between the formants and in fibers with high BFs. If the power ratio is computed separately for each harmonic, then a tonotopic distribution of each component is observed, in which response to a particular harmonic is centered on its place in the population (Schilling, 1995). As stimulus level increases $PR(\text{trough})$ does not increase; in fact the total power at the trough frequencies decreases slightly at the highest level (Fig. 14A), mainly in the vicinity of the formants. Apparently, the re-

sponses to trough frequencies are suppressed by the formants, so that phase locking to the trough frequencies is very low in fibers with BFs near the formants.

In impaired fibers, the overall level of $PR(\text{trough})$ is roughly the same as in normal fibers, but the distribution is much broader. In particular, suppression of phase locking to the trough frequencies is not observed at the formant BF places. The results in Fig. 14C–F show that the broadband responses seen in the examples for the impaired fibers in Figs. 9 and 10 are typical of the population.

III. DISCUSSION

We have examined the permanent deficits in the auditory-nerve code for the vowel / ϵ / that are produced by acoustic trauma. The aims of this study were to gain insight into the neural code for speech after cochlear damage and to understand better the neurophysiological basis for deficits experienced by listeners with a sensorineural hearing loss. A significant decrease in the quality of the phase locked response to the vowel was found, reflected mainly in a broadband response to the vowel and a loss of synchrony capture by the formant frequencies.

A. Nature of the lesion

We did not do histological examinations of the cochleas in our impaired animals, so the nature of the hair cell lesions cannot be determined directly. However, our paradigm for inducing acoustic trauma closely followed that of Liberman and co-workers (Liberman and Kiang, 1978; Liberman and Beil, 1979; Liberman and Mulroy, 1982; Liberman and Dodds, 1984), who have previously made a thorough study of the correlation between FTC changes and damage to inner and outer hair cells. In cats, a 2-h exposure to narrow-band noise presented at 110–115 dB SPL gives rise to an affected band of hair cells either centered on the exposure frequency or a region of damage which spreads preferentially towards the base of the cochlea. Damage to inner versus outer hair cells and their stereocilia gives rise to different types of FTC abnormalities. In our study, we make use of this established relationship to infer the cochlear pathology from FTC data.

Damage to or loss of only outer hair cells has been shown to result in a loss of the sensitive tip region of the FTC, broadened tuning, and in many cases a hypersensitivity in the tail region of the FTC (Dallos and Harris, 1978; Robertson and Johnstone, 1979; Robertson, 1982; Liberman and Dodds, 1984). We found wide ranges of BFs in which all fibers showed a loss of the tip regions and broadened tuning, which are consistent with outer hair cell damage. However, we found few fibers with hypersensitive tails. The impaired animals are likely to have also suffered substantial damage to inner hair cells, because fibers with BFs near the exposure frequency are under-represented in the impaired pool (Fig. 4C), relative to the normal pool (Fig. 3C). This finding argues for inner hair cell damage, followed by silencing and perhaps degeneration of auditory-nerve fibers. The FTCs of fibers with BFs near the exposure frequency are similar to those from cochlear regions with damage to both the inner and outer hair cells (Liberman and Mulroy, 1982). The pau-

city of hypersensitive tuning curves in our preparations is also consistent with damage to inner hair cell stereocilia, which has been shown to result in a sensitivity loss at all frequencies (Liberman and Dodds, 1984).

B. Changes in phase locking

A principal result of this paper is the degraded phase locked representation of the stimulus in impaired animals. It is necessary to consider whether this degradation is partly due to changes in the ability of auditory-nerve fibers to phase lock after acoustic trauma. A decrease in phase locking to tones has been reported in chinchillas following kanamycin ototoxicity (Woolf *et al.*, 1981). However the ability of fibers to support the phase locking necessary for reverse correlation studies is apparently not affected in guinea pigs (Harrison and Evans, 1982). Although some questions remain that are related to the behavior of maximum synchrony as opposed to $S_{50\text{ dB}}$, it is evident that a systematic loss of phase locking did not occur in our impaired animals. In any case, we cannot account for the loss of synchrony capture by damage to the forward transduction pathway and argue below that the deficit affecting phase locking in response to the vowel resides in the reverse transduction pathway.

C. Correlation between degraded temporal representation of / ϵ / and diminished two-tone suppression

Nonlinear suppression phenomena clearly participate in the generation of responses to speech stimuli in the normal cochlea. For example, the maintenance of a good temporal representation of spectral shape as sound level increases depends on suppression of the growth of phase locking to stimulus harmonics in the trough between $F1$ and $F2$ (Young and Sachs, 1979). Similarly, the growth of discharge rate at BFs equal to and above $F2$ is limited by two-tone rate suppression (Sachs and Young, 1980). Such a limitation is not predicted by the FTCs of auditory nerve fibers; models which mimic FTC shape but lack explicit suppression have discharge rates that increase monotonically up to saturation (Payton, 1988). In auditory-nerve fibers, the wide spread of phase locking to $F1$ along the tonotopic axis in normal animals suggests that the suppressor is the energy in the stimulus at $F1$.

Suppression is also evident in the synchrony capture phenomenon seen in normal fibers with BFs near the formant frequencies. At low sound levels (e.g., Fig. 9C), fibers respond to several harmonics in the stimulus, as would be predicted by the set of harmonics falling within the fiber's FTC. However, as stimulus level increases (e.g., Fig. 9A and B), the response is captured by the strongest harmonic, in this case a formant frequency, and responses to the other stimulus components are suppressed.

We have demonstrated a progressive loss of two-tone rate suppression as the threshold shift in a region of the cochlea increases (Fig. 6). This result is consistent with earlier studies of cochleas subsequent to acoustic trauma (Schmiedt *et al.*, 1980, 1990; Salvi *et al.*, 1982), kanamycin induced outer hair cell lesions (Dallos *et al.*, 1980; Schmiedt *et al.*,

1980) and perilymph removal (Robertson, 1976). It is tempting to speculate that there is a connection between this weakening of rate suppression and the loss of synchrony capture in our data, under the assumption that synchrony capture involves suppression of other response harmonics by the capturing harmonic. However, the interpretation of the loss of synchrony capture is not so clear, because of the significant change in tuning that also occurs with acoustic trauma.

The effects of broad tuning on response synchrony have been investigated by Geisler (1989), who used linear filters similar in shape to the FTCs of the impaired fibers in Fig. 8 to model tuning in a damaged cochlea. The outputs of the filters were applied to a synaptic model for the inner hair cell/afferent-fiber complex. The only nonlinearities in the model were those in this synapse. Simulation of afferent fibers' firing probabilities to a vowel in the presence of weak noise (+30 dB signal-to-noise ratio) showed abnormally strong synchrony to $F1$, and in several cases (e.g., Geisler's Figs. 4e, 5e, 6e, and 7e) it is clear that the model fibers are responding to more than one formant component of the stimulus. These results suggest that the expansion of $F1$ phase locking observed in our data can be at least partially accounted for by the abnormal tuning curve shapes.

In fact, a loss of synchrony capture similar to that seen in Fig. 9A–C can be observed in a cochlear model which ordinarily displays synchrony capture (Carney, 1993) by broadening the bandwidth of the filter at the input stage of the model (unpublished results). The change in cochlear filter width and the loss of two-tone suppression that are seen with acoustic trauma are both thought to derive from loss of outer hair cells (Patuzzi *et al.*, 1989). Thus it is not clear to what extent it is profitable to attempt to distinguish between these two potential mechanisms for the loss of synchrony capture.

D. Implications for hearing impairment

The impairments in the temporal representation of vowels clearly indicate a loss of specificity in the response of auditory-nerve fibers to complex stimuli. The broadband nature of the responses in impaired fibers predicts that impaired listeners should experience difficulty in identifying the spectral peaks present in speech stimuli and in separating out background stimuli. This prediction is in keeping with studies of listeners with sensorineural hearing loss which show that the performance obtained in an ideal listening environment, i.e., a loud speaker in a quiet surrounding, is degraded faster than normal as competing sound sources are added (Plomp, 1978; Glasberg and Moore, 1989; Smoorenburg, 1992).

Palmer and Moorjani (1993) have described deficits in the representation of double vowels in guinea pig auditory nerve after kanamycin ototoxicity. In contrast to our results, their single-vowel population responses were quite similar to those from normal animals. They attribute this primarily to the inclusion of fibers with relatively small threshold shifts. However, the population response to double vowels showed increased synchrony to the lowest fundamental frequency of the double vowel stimulus, accompanied by a shift in the synchrony away from the higher frequency components. This is consistent with the upward spread of synchrony from

the $F2$ place to the $F3$ place, the concomitant loss of synchrony to $F3$ (Fig. 13D–F), and the weakening of synchrony to $F2$ (Fig. 11D–F) observed in the data reported here. Such defects clearly provide a biophysical basis for the difficulties experienced by hearing impaired listeners in the presence of speech-like background noise.

The analyses in this paper have centered on phase locking, i.e., on the temporal representation of speech. There is no evidence for or against the proposition that the brain makes use of information encoded in phase locking for extraction of spectral information. Nevertheless, analysis of temporal response patterns is an appropriate and powerful tool for the study of information representation in the auditory nerve because it allows us to identify the stimulus components to which a neuron is responding. That is, temporal analysis reveals much about the nature of impairment and provides guidance as to the problems that need to be solved in order to compensate for the impairment.

The acoustic trauma paradigm described here was designed to induce a hearing deficit that can be used to assess the effects of various signal processing paradigms for hearing aids. The simplest such aid is amplification which restores audibility to the frequency regions with a threshold loss, without compensating for regions which possess normal thresholds. Our results clearly show that making the stimulus louder restores audibility, but in the form of a broadband response. Without restoring the normal response pattern, the vowel's spectral features remain blurred as compared to a normal ear. Restoration of a normal temporal-place response pattern represents a significant challenge in the design of an auditory prosthesis.

ACKNOWLEDGMENTS

The authors thank Dr. Murray Sachs and Dr. Brad May for helpful comments at various stages of this work. Technical assistance of Cynthia Aleszczyk, Ron Atkinson, and Phyllis Taylor is also appreciated. This work was supported by a grant from The Johns Hopkins Center for Language and Speech Processing and NIDCD Grant No. DC00109.

Abbas, P. J., and Sachs, M. B. (1976). "Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship," *J. Acoust. Soc. Am.* **59**, 112–122.

Cai, Y., and Geisler, C. D. (1996). "Suppression in auditory-nerve fibers of cats using low-side suppressors. II. Effect of spontaneous rates," *Hear. Res.* **96**, 113–125.

Carney, L. H. (1993). "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am.* **93**, 401–417.

Costalupes, J. A., Young, E. D., and Gibson, D. J. (1984). "Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat," *J. Neurophysiol.* **51**, 1326–1344.

Dallos, P., and Harris, D. (1978). "Properties of auditory nerve responses in absence of outer hair cells," *J. Neurophysiol.* **41**, 365–383.

Dallos, P., Harris, D. M., Relkin, E., and Cheatham, M. A. (1980). "Two-tone suppression and intermodulation distortion in the cochlea: Effect of outer hair cell lesions," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft U.P., Delft), pp. 242–252.

Delgutte, B., and Kiang, N. Y. (1984). "Speech coding in the auditory nerve: I. Vowel-like sounds," *J. Acoust. Soc. Am.* **75**, 866–878.

Evans, E. F., and Wilson, J. P. (1973). "The frequency selectivity of the cochlea," in *Basic Mechanisms in Hearing*, edited by A. R. Møller (Academic, New York), pp. 519–554.

Franck, K. R. (1994). "Auditory-nerve responses to speech sounds in partially deafened cats," Master of Science in Engineering thesis, The Johns Hopkins University.

Geisler, C. D. (1989). "The responses of models of 'high-spontaneous' auditory-nerve fibers in a damaged cochlea to speech syllables in noise," *J. Acoust. Soc. Am.* **86**, 2192–2205.

Glasberg, B. R., and Moore, B. C. (1989). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," *Scand. Audiol. Suppl.* **32**, 1–25.

Guinan, J. J., and Peake, W. T. (1967). "Middle-ear characteristics of anesthetized cats," *J. Acoust. Soc. Am.* **41**, 1237–1261.

Harrison, R. V., and Evans, E. F. (1982). "Reverse correlation study of cochlear filtering in normal and pathological guinea pig ears," *Hear. Res.* **6**, 303–314.

Johnson, D. H. (1980). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," *J. Acoust. Soc. Am.* **68**, 1115–1122.

Kiang, N. Y., Moxon, E. C., and Levine, R. A. (1970). "Auditory-nerve activity in cats with normal and abnormal cochleas," in *Sensorineural Hearing Loss*, edited by G. E. W. Wolstenholme and J. Knight (Churchill, London), pp. 241–273.

Kiang, N. Y., Liberman, M. C., and Levine, R. A. (1976). "Auditory-nerve activity in cats exposed to ototoxic drugs and high-intensity sounds," *Ann. Otol. Rhinol. Laryngol.* **85**, 752–768.

Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.

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. (1984). "Single-neuron labeling and chronic cochlear pathology. I. Threshold shift and characteristic-frequency shift," *Hear. Res.* **16**, 33–41.

Liberman, M. C., and Beil, D. G. (1979). "Hair cell condition and auditory nerve response in normal and noise-damaged cochleas," *Acta. Oto-Laryngol. (Stockh.)* **88** (3-4), 161–176.

Liberman, M. C., and Dodds, L. W. (1984). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," *Hear. Res.* **16**, 55–74.

Liberman, M. C., and Kiang, N. Y. (1978). "Acoustic trauma in cats: Cochlear pathology and auditory-nerve activity," *Acta. Oto-Laryngol. Suppl. (Stockh.)* **358**, 1–63.

Liberman, M. C., and Mulroy, M. J. (1982). "Acute and chronic effects of acoustic trauma: Cochlear pathology and auditory nerve pathophysiology," in *New Perspectives on Noise-Induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. Salvi (Raven, New York), pp. 105–135.

Miller, J. D., Watson, C. S., and Covell, W. P. (1963). "Deafening effects of noise on the cat," *Acta. Oto-Laryngol. Suppl. (Stockh.)* **176**, 1–81.

Palmer, A. R. (1990). "The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers," *J. Acoust. Soc. Am.* **88**, 1412–1426.

Palmer, A. R., and Evans, E. F. (1982). "Intensity coding in the auditory periphery of the cat: Responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise," *Hear. Res.* **7**, 305–323.

Palmer, A. R., and Moorjani, P. A. (1993). "Responses to speech signals in the normal and pathological peripheral auditory system," *Prog. Brain Res.* **97**, 107–115.

Palmer, A. R., Winter, I. M., and Darwin, C. J. (1986). "The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons," *J. Acoust. Soc. Am.* **79**, 100–113.

Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (1989). "Outer hair cell receptor current and sensorineural hearing loss," *Hear. Res.* **42**, 47–72.

Payton, K. L. (1988). "Vowel processing by a model of the auditory periphery: A comparison to eight-nerve responses," *J. Acoust. Soc. Am.* **83**, 145–162.

Plomp, R. (1978). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," *J. Acoust. Soc. Am.* **63**, 533–549.

- Prijs, V. F. (1989). "Lower boundaries of two-tone suppression regions in the guinea pig," *Hear. Res.* **42**, 73–81.
- Robertson, D. (1976). "Correspondence between sharp tuning and two-tone inhibition in primary auditory neurones," *Nature* **259**, 477–478.
- Robertson, D. (1982). "Effects of acoustic trauma on stereocilia structure and spiral ganglion cell tuning properties in the guinea pig cochlea," *Hear. Res.* **7**, 55–74.
- Robertson, D., and Johnstone, B. M. (1979). "Aberrant tonotopic organization in the inner ear damaged by kanamycin," *J. Acoust. Soc. Am.* **66**, 466–469.
- Sachs, M. B., and Young, E. D. (1980). "Effects of nonlinearities on speech encoding in the auditory nerve," *J. Acoust. Soc. Am.* **68**, 858–875.
- Salvi, R., Perry, J., Hamernik, R. P., and Henderson, D. (1982). "Relationships between cochlear pathologies and auditory nerve and behavioral responses following acoustic trauma," in *New Perspectives on Noise-Induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. Salvi (Raven, New York), pp. 165–188.
- Schilling, J. R. (1995). "The effects of acoustic trauma and hearing-aid processing on the encoding of vowels in the auditory nerve," Master of Science in Engineering thesis, The Johns Hopkins University.
- Schmiedt, R. A., Mills, J. H., and Adams, J. C. (1990). "Tuning and suppression in auditory nerve fibers of aged gerbils raised in quiet or noise," *Hear. Res.* **45**, 221–236.
- Schmiedt, R. A., Zwislocki, J. J., and Hamernik, R. P. (1980). "Effects of hair cell lesions on responses of cochlear nerve fibers. I. Lesions, tuning curves, two-tone inhibition, and responses to trapezoidal-wave patterns," *J. Neurophysiol.* **43**, 1367–1389.
- Sinex, D. G., and Geisler, C. D. (1983). "Responses of auditory-nerve fibers to consonant-vowel syllables," *J. Acoust. Soc. Am.* **73**, 602–615.
- 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.
- Sokolich, W. G. (1977). "Improved acoustic system for auditory research," *J. Acoust. Soc. Am. Suppl. 1* **62**, S12.
- Wiener, F. M., and Ross, D. A. (1946). "The pressure distribution in the auditory canal in a progressive sound field," *J. Acoust. Soc. Am.* **18**, 401–408.
- Woolf, N. K., Ryan, A. F., and Bone, R. C. (1981). "Neural phase-locking properties in the absence of cochlear outer hair cells," *Hear. Res.* **4**, 335–346.
- Young, E. D., Miller, R. L., Wong, J. C., and Schilling, J. R. (1996). "Changes in phase locking to complex stimuli produced by noise-induced hearing loss," in *Diversity in Auditory Mechanics* (World Scientific, Singapore, in press), pp. 595–601.
- Young, E. D., and Sachs, M. B. (1979). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," *J. Acoust. Soc. Am.* **66**, 1381–1403.

The modulated–unmodulated difference: Effects of signal frequency and masker modulation depth^{a)}

Sid P. Bacon^{b)} and Jungmee Lee

Psychoacoustics Laboratory, Department of Speech and Hearing Science, P.O. Box 871908, Arizona State University, Tempe, Arizona 85287-1908

(Received 16 July 1996; revised 2 December 1996; accepted 31 January 1997)

The masked threshold for a signal is oftentimes lower when the masker is modulated than when it is unmodulated. The difference in masked thresholds is referred to as the modulated–unmodulated difference, or MUD. The purpose of the present study was to follow up on the results of a previous study [Bacon *et al.*, *J. Acoust. Soc. Am.* **101**, 1600–1610 (1997)] which showed that the MUD is larger for high than for low signal frequencies, both when the masker is no wider than a critical band (and the processing is solely within channel) and when it is broadband (and the processing may be both within and across channel). The present results indicate that the effects of signal frequency primarily exist only when the modulated masker is modulated at a depth greater than about 0.75, and that at these large depths, thresholds in the presence of the modulated masker are governed largely by forward masking. By far, the effect of signal frequency is larger with the broadband masker than with the critical-band masker, suggesting that there may be an across-channel process whose contribution is greater at high than at low signal frequencies. It is argued here that this across-channel process may be related to psychophysical suppression. © 1997 Acoustical Society of America. [S0001-4966(97)00706-6]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [WJ]

INTRODUCTION

We recently completed a series of experiments designed to compare the masking produced by a modulated masker with that produced by an unmodulated masker (Bacon *et al.*, 1997). The difference in masking is referred to as the modulated–unmodulated difference, or MUD (Carlyon *et al.*, 1989). In our experiments, the MUD was obtained with a masker whose bandwidth was approximately the width of a critical band, or equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990), as well as with a broadband (BB) masker. The MUD obtained with an ERB masker (MUD_{ERB}) may provide an estimate of within-channel processing, whereas that obtained with a BB masker (MUD_{BB}) may represent both within- and across-channel processing. An estimate of the across-channel processing can be obtained by $MUD_{BB}-MUD_{ERB}$ (Carlyon *et al.*, 1989); this is also a measure of comodulation masking release (CMR) (Hall *et al.*, 1984).

In our previous study, we found, among other things, that the MUD_{ERB} and the MUD_{BB} are larger at high than at low signal frequencies. Moreover, the effect of signal frequency was larger for the MUD_{BB} , suggesting that there may be both a within-channel and an across-channel effect of signal frequency. The main objective of the present study was to follow up on that finding, and in particular to conduct experiments designed to provide new insights into the pos-

sible mechanisms underlying the effect of signal frequency on the MUD.

I. GENERAL METHODS

A. Apparatus and stimuli

The sinusoidal signals were digitally generated and produced (TDT DA1) at a 20 000-Hz sampling rate and subsequently low-pass filtered at 8000 Hz (Kemo VBF 25.01; 135 dB/oct). The noise masker was created by multiplying a sinusoid (HP 8904A synthesizer) whose frequency was equal to the signal frequency by a low-pass filtered (Kemo VBF 25.01) random noise (GenRad 1381). The bandwidth of the multiplied noise was twice the cut-off frequency of the low-pass filter.

Unless otherwise stated, the duration of the signal and masker was 500 ms, including the 20-ms \cos^2 rise–fall times. The bandpass noise was gated via a digital gating function (TDT DA1; 20 000-Hz sampling rate) that was low-pass filtered at 8000 Hz (Kemo VBF 25.01) prior to multiplication (Analog Devices 435 K). The gating function determined whether the masker was unmodulated or sinusoidally amplitude modulated (SAM).¹ The sinusoidal modulator had a starting phase of 270°, so that the SAM masker began at its minimum.

The levels of the signal and masker were adjusted via programmable attenuators (Wilsons PATT). Signal level was varied adaptively whereas masker level was fixed throughout a block of trials. The masker level for a given experiment is noted in the context of that experiment. The signal and masker were added together prior to being transduced by a TDH-49P headphone mounted in an MX/51 cushion.

^{a)}Portions of this research were presented at the 129th meeting of the Acoustical Society of America [S. P. Bacon and J. Lee, “Temporal resolution and CMR can depend upon frequency,” *J. Acoust. Soc. Am.* **97**, 3273(A) (1995)].

^{b)}Corresponding author: Electronic mail: spb@asu.edu

B. Procedure

Thresholds were measured using an adaptive, two-interval, forced-choice (2IFC) procedure that tracked 79.4% correct detection (Levitt, 1971). Each run consisted of a block of 60 trials. The initial step size of 4 dB was reduced to 2 dB after the first two reversals. The first two (or three, if the total number of reversals was odd) reversal points were discarded, and the threshold estimate for a given block was based on the average of the signal level at the remaining reversal points. On a few occasions, runs were discarded because there were fewer than six reversals in the mean or the standard deviation of the given threshold estimate was greater than 5 dB. Final threshold values for a particular condition are based on the average of at least three runs. If the standard deviation of that average was greater than 3 dB, an additional estimate was obtained and included in the average. This continued until the standard deviation was less than 3 dB or until six estimates were obtained. Only rarely (10/336 conditions) were more than three estimates needed, and the standard deviation was less than 3 dB in over 98% (331/336) of the conditions.

C. Subjects

Two females and one male served as subjects. They ranged in age from 22 to 29 years, with a mean age of 25 years. Their absolute thresholds were not worse than 15 dB HL (ANSI 1989) at octave frequencies from 250 to 4000 Hz. One subject was author JL; the other two subjects were paid an hourly wage for their participation.

II. EXPERIMENT 1: EFFECTS OF SIGNAL FREQUENCY

A. Rationale and conditions

In most of the experiments from our previous study (Bacon *et al.*, 1997), subjects were tested at only two signal frequencies. The purpose of experiment 1 was to extend those results to include a large range and number of signal frequencies, in order to evaluate in detail the effects of signal frequency on the MUD.

The signal frequency was 250, 500, 1000, 2000, or 4000 Hz. The bandwidth of the masker was one ERB or BB (10 kHz). The bandwidths (in Hz) for the ERB masker at the various signal frequencies were 60 at 250 Hz; 80 at 500 Hz; 140 at 1000 Hz; 240 at 2000 Hz; and 460 at 4000 Hz. The spectrum level of the masker was 40 dB SPL. The masker was either unmodulated ($m=0.0$) or SAM at a depth (m) of 1.0. The modulation rate of the masker was 4, 8, or 16 Hz.

B. Results and discussion

The results generally were similar across the three subjects, and thus only their group mean data are shown. In Fig. 1, the results are plotted both in terms of signal threshold (top panels) and in terms of the MUD (bottom panels). The results for the ERB masker are in the left panels, and those for the BB masker are in the right panels. First consider the thresholds obtained in the presence of the ERB masker (top left panel). When the masker was unmodulated (unfilled triangles), threshold increased by about 6 dB as the signal fre-

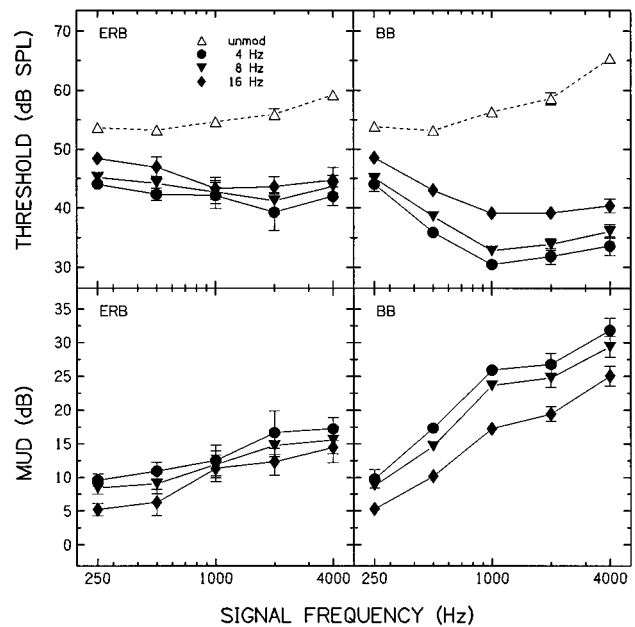


FIG. 1. The top row shows thresholds as a function of signal frequency in the presence of the ERB (left panel) and BB (right panel) masker; the results have been averaged across the three subjects. Unfilled triangles represent thresholds when the masker was unmodulated; filled symbols represent thresholds when the masker was modulated at a depth (m) of 1.0. Error bars indicate plus and minus one standard error. The bottom row shows the modulated-unmodulated difference (MUD) calculated from the thresholds in the top row. Error bars indicate plus and minus one standard error, calculated as: $s.e.^2_{MUD} = s.e.^2_{unmod} + s.e.^2_{mod}$.

quency increased from 250 to 4000 Hz; this increase is probably due to the fact that the overall masker level increased (by about 9 dB) as the signal frequency (and hence masker bandwidth) increased. When the masker was modulated (filled symbols), threshold tended to decrease by about 3 to 4 dB as the signal frequency increased from 250 to about 1000 or 2000 Hz, but then it remained roughly constant or even increased slightly with further increases in frequency. The difference between the threshold in the presence of the unmodulated masker and that in the presence of the modulated masker defines the MUD. The MUD_{ERB} (bottom left panel) increased with increasing signal frequency, from a value of about 8 dB at 250 Hz to one of about 16 dB at 4000 Hz. This 8 dB increase in MUD_{ERB} with increasing signal frequency reflects the changes in threshold as a function of frequency observed for both the unmodulated and modulated maskers.

The effects observed with the ERB masker were also observed with the BB masker (right panels), but to a greater extent. First consider the threshold data. For the unmodulated masker (unfilled symbols), threshold increased by almost 12 dB as signal frequency increased from 250 to 4000 Hz. This increase is 6 dB larger than that observed with the ERB masker, and it is largely due to the especially high threshold at 4000 Hz (which is 6 dB higher than the corresponding threshold in the presence of the ERB masker). Although one would expect the threshold to increase with increasing signal frequency with the BB masker (e.g., Hawkins and Stevens, 1950), given that the overall level in the critical band increases with signal frequency, it is unclear why the threshold in the presence of the BB masker is so much higher

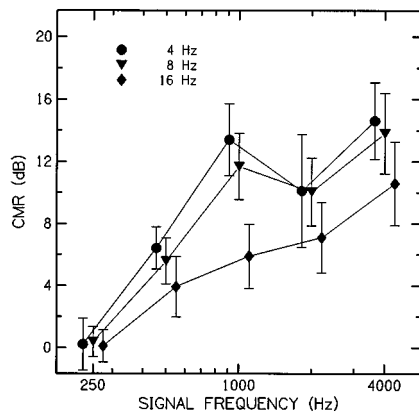


FIG. 2. Comodulation masking release (CMR) obtained from the values in Fig. 1. CMR is given by $MUD_{BB} - MUD_{ERB}$. Data points have been offset horizontally for clarity. Error bars indicate plus and minus one standard error, calculated as: $s.e.^2_{CMR} = s.e.^2_{MUD(ERB)} + s.e.^2_{MUD(BB)}$.

than it is in the presence of the ERB masker. Some of the difference may be due to our subjects having larger-than-normal critical bands at high frequencies, but it seems very unlikely that this can account for the entire effect, as it would require that their critical bands were four times larger than normal. In the presence of the modulated masker (filled symbols), threshold declined by about 12 dB as signal frequency increased from 250 to 1000 Hz, but then remained roughly constant thereafter. The MUD_{BB} (bottom right panel) increased monotonically with signal frequency. The approximately 14 dB increase in the MUD_{BB} as signal frequency increased from 250 to 1000 Hz was primarily due to the decline in threshold in the presence of the modulated masker, whereas the nearly 6-dB increase as signal frequency was increased further was due to the increase in threshold in the presence of the unmodulated masker.

The effect of signal frequency on the MUD was clearly larger for the BB than for the ERB masker, suggesting that there is both a within-channel and an across-channel component that is influenced by signal frequency. An estimate of the size of this across-channel component is given by $MUD_{BB} - MUD_{ERB}$. As noted in the Introduction, this is also a measure of CMR. The values of CMR from the present data are shown in Fig. 2. The amount of CMR increased from 0 dB at 250 Hz to about 10–14 dB at 4000 Hz. At the two lower modulation rates (4 and 8 Hz), CMR reached an asymptote at 1000 Hz, whereas at the higher rate (16 Hz) CMR tended to increase with increasing frequency throughout the range of frequencies tested here. For a given signal frequency, the amount of CMR was generally smaller at a modulation rate of 16 Hz than it was at 4 or 8 Hz; there was essentially no difference in the amount of CMR at the two lower rates. This effect of modulation rate reflects the fact that the MUD—particularly for the BB masker—was smaller at 16 Hz than it was at 4 or 8 Hz (Fig. 1). We have observed a similar effect previously, although the effects of rate on the MUD_{ERB} , MUD_{BB} , and CMR in the literature have been rather mixed, possibly due to differences in modulator type (see Bacon *et al.*, 1997, for a review).

The effects of signal frequency on CMR observed here

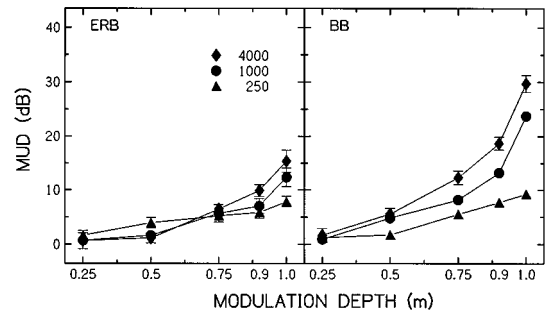


FIG. 3. MUD values as a function of the modulation depth of the modulated masker; the masker rate was 8 Hz. The results have been averaged across the three subjects. Error bars indicate plus and minus one standard error (see caption for Fig. 1).

differ from those of Schooneveldt and Moore (1987), who found little effect of signal frequency (for frequencies ranging from 250 to 8000 Hz) on what they called the “true” CMR. Our results also differ from those of Haggard *et al.* (1990), who reported no difference in the magnitude of CMR for signal frequencies from 500 to 4000 Hz, once their maskers were scaled in terms of the critical bandwidth at the signal frequency. It is unclear why the present results differ from previous results, although it may be related to differences in the modulation across studies. The modulation used in the two previous studies was the aperiodic modulation inherent in narrow bands of noise, whereas the modulation used here was well-defined and periodic. Further, the equivalent modulation rates in the two previous studies were about 16 Hz (Schooneveldt and Moore, 1987) or 32 Hz (Haggard *et al.*, 1990), which may have resulted in smaller effects of signal frequency, given the trend for the frequency effect to decrease at higher modulation rates (Fig. 2).

III. EXPERIMENT 2: EFFECTS OF MASKER DEPTH

A. Rationale and conditions

In one of the experiments in Bacon *et al.* (1997), the MUD was obtained as a function of masker bandwidth when the modulated masker was modulated at depth of 0.75 or 1.0. The signal frequency was 250, 2000, or 4000 Hz. The effect of signal frequency was considerably reduced when the masker was modulated at a depth of 0.75 (see their Fig. 4). The purpose of experiment 2 was to extend those results to a wide range of masker depths. The signal frequency was 250, 1000, or 4000 Hz. The bandwidth of the masker was one ERB or BB. The masker was unmodulated or SAM at a rate of 8 Hz and a depth of 0.25, 0.5, 0.75, 0.9, or 1.0. The spectrum level of the masker was 40 dB SPL.

B. Results and discussion

The results were similar across the three subjects, and thus only their mean data are shown. In Fig. 3, the MUD is plotted as a function of the modulation depth of the masker for both the ERB (left panel) and BB masker (right panel); signal frequency is the parameter. For both masker bandwidths, the MUD tended to increase as masker depth increased, particularly for the two higher signal frequencies.

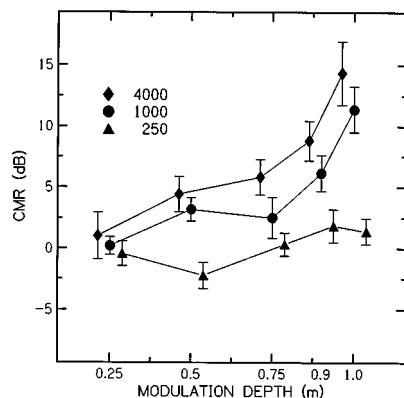


FIG. 4. CMR values from the data in Fig. 3. Data points have been offset horizontally for clarity. Error bars indicate plus and minus one standard error (see caption for Fig. 2).

As the modulation depth of the BB masker increased from 0.25 to 1.0, the MUD increased from about 0 to 10 dB for the 250-Hz signal (triangles), 0 to 25 dB for the 1000-Hz signal (circles), and 0 to 30 dB for the 4000-Hz signal (diamonds). The effect of modulation depth was considerably smaller for the ERB masker, where the MUD increased from about 0 dB to anywhere from 8 to 15 dB, depending upon the signal frequency.

It is important to note that an effect of signal frequency on the MUD is observed only at the larger modulation depths. The effect of signal frequency is essentially eliminated at a depth of 0.75 for the ERB masker and 0.50 for the BB masker. The MUD values at those depths are considerably smaller than those at the larger depths (indeed, the MUD converges to a value close to 0 dB at a depth of 0.25), but the frequency effect largely disappears before the MUD values reach 0 dB.

The amount of CMR is shown in Fig. 4. Except for the 250-Hz signal (triangles), where CMR was negligible at all masker depths, the amount of CMR generally increased with increasing masker depth, as has been seen, at least to some extent, in experiments on CMR with SAM maskers consisting of multiple tone or narrow-band noise carriers (Grose and Hall, 1989, and Eddins and Wright, 1994; Hicks and Bacon, 1995). For all three signal frequencies, there was essentially no CMR at a masker depth of 0.25. For the two higher frequencies, the amount of CMR was about 3–5 dB for masker depths of 0.5 and 0.75, but it increased abruptly to about 12 to 15 dB at a depth of 1.0. This rather dramatic increase in CMR when the masker depth increased beyond 0.75 suggests the possibility of a change in (or an additional contribution to) the processing underlying CMR at these larger depths, a possibility to which we now turn.

IV. INTERIM DISCUSSION

We argued previously that thresholds in the presence of a masker modulated at a depth of 1.0 were governed primarily by forward masking (Bacon *et al.*, 1997). In particular, we suggested that subjects listen in the dips of the modulated masker, and that the masking is dominated by forward masking from the preceding peak in the masker's envelope. Con-

sistent with that possibility were the growth-of-masking functions, which had a slope of one when the masker was unmodulated, but a slope much less than that (as would be expected for forward masking) when the masker was modulated at a depth of 1.0. When the masker was modulated at a depth of 0.75, the masking functions had a slope near one, suggesting that thresholds in the presence of that masker were governed more by simultaneous masking in a dip than by forward masking from a preceding peak.

The importance of determining to what extent thresholds in the presence of a modulated masker are governed by forward masking is underscored by the fact that the processing underlying both the MUD_{ERB} and MUD_{BB} , and hence the CMR in this paradigm, may vary depending upon the modulation depth of the masker. In particular, if forward masking dominates at large modulation depths, then it is possible—indeed, likely—that processes revealed in forward masking will contribute to the overall magnitude of the MUD and the resulting CMR at these depths, but that they will not contribute at smaller modulation depths. The rather abrupt increase in the MUD (Fig. 3) and the CMR (Fig. 4) when the masker depth increased from 0.75 to 0.9 (and then to 1.0) is certainly consistent with an additional process (or additional processes) contributing at the larger depths. Importantly, this means that the signal frequency effect observed here and in Bacon *et al.* (1997) may result from, or at least be closely related to, processes revealed only in forward masking, since the effect is largely eliminated for masker depths of 0.75 and below. Experiment 3 was designed to evaluate the extent to which forward masking might play a role in the MUD experiments reported here and elsewhere.

V. EXPERIMENT 3: GROWTH OF MASKING

A. Rationale and conditions

Although the growth-of-masking functions in Bacon *et al.* (1997) suggest that forward masking dominates when the modulated masker is modulated at a depth of 1.0, and that simultaneous masking dominates when it is modulated at 0.75, a better test of those possibilities would be to compare the masking produced by a modulated masker to that produced by an unmodulated masker whose level was equal to the level of a dip in the modulated masker. This was the purpose of experiment 3.

Thresholds were measured as a function of the level of an unmodulated masker whose bandwidth was one ERB or BB. The spectrum level of the masker was 5, 20, 29, 34 or 37 dB SPL. These levels correspond to the levels in the dip of a 40-dB masker modulated at a rate of 8 Hz and a depth of 1.0, 0.9, 0.75, 0.5, or 0.25. The levels were determined by centering a 20-ms Hanning window (HP 3561 signal analyzer) in the dip of a masker envelope. The signal frequency was 250 or 4000 Hz. The signal had a duration of 30 ms (10-ms \cos^2 rise–fall times with 10-ms steady state), as opposed to the duration of 500 ms that was used in experiments 1 and 2, in an attempt to equate the “effective” signal duration for conditions where subjects listen in a dip in the envelope of a modulated masker modulated at a rate of 8 Hz.

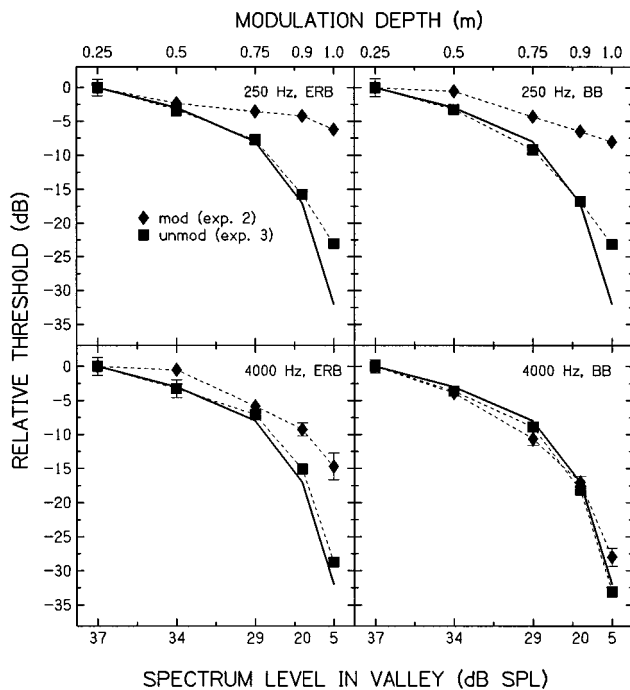


FIG. 5. The squares represent thresholds for a 30-ms signal obtained in the presence of an unmodulated simultaneous masker, whose spectrum level ranged from 37 to 5 dB SPL (results from experiment 3). Thresholds are plotted relative to the threshold obtained when the masker was 37 dB SPL. The solid line without symbols represents the relative level of the masker in experiment 3. The diamonds represent thresholds for a 500-ms signal obtained in the presence of a modulated masker, whose modulation depth ranged from 0.25 to 1.0 (results from experiment 2). Thresholds are plotted relative to the threshold obtained when the masker was modulated at a depth of 0.25. The results from both experiments have been averaged across the three subjects, and error bars represent plus and minus one standard error.

The signal was temporally centered within the 500-ms masker.

B. Results and discussion

The results were similar across the three subjects, and thus only their mean data are shown in Fig. 5. The growth-of-masking data from the present experiment are shown as squares (referred to the bottom axis), and the results from experiment 2 on the effects of masker depth are shown as diamonds (referred to the top axis). Within a panel, the thresholds are referenced to the threshold obtained at a masker level of 37 dB (squares) or at a masker depth of 0.25 (diamonds). The results are plotted in terms of relative thresholds, rather than the thresholds themselves, because of the difficulty in equating the effective signal duration between the two experiments.² Also shown in Fig. 5 as a solid line without symbols are the relative masker levels. One would expect the results from the present experiment to follow those levels closely. To the extent that the thresholds in the presence of the modulated masker are governed by simultaneous masking in the dips, the relative thresholds from experiment 2 should also follow those levels.

Ignoring, for the moment, the fact that the results from experiment 2 and 3 overlap for a signal frequency of 4000 Hz in the presence of a BB masker (bottom right panel), the

results in this figure can be summarized as follows. The relative thresholds in the present experiment with an unmodulated masker overlap the relative masker levels, as would be expected based on growth-of-masking functions for noise maskers (e.g., Hawkins and Stevens, 1950). The exception to this is at the lowest masker level for the 250-Hz signal (top panels), where the masked threshold was limited by the quiet threshold. The relative thresholds from experiment 2 with a modulated masker (diamonds), however, overlap or nearly overlap the relative thresholds from the present experiment only at modulation depths of about 0.75 and below (again, ignoring the bottom right panel). These results strongly suggest that thresholds in the presence of a masker modulated at relatively low depths are governed by simultaneous masking in the dips. At modulation depths above about 0.75, the thresholds from the two experiments diverge, suggesting that the thresholds in the presence of the modulated masker are no longer governed by simultaneous masking in the dips, but instead may be influenced significantly by forward masking.

Now consider the results in the bottom right panel, (BB masker and 4000-Hz signal). At first glance, the finding that the relative thresholds from the two experiments overlap at all masker depths may seem to indicate that, for this condition alone, thresholds in the presence of the modulated masker are determined by simultaneous masking in the dips at all masker depths, not just those below about 0.75. It seems unlikely, however, that this would be the case *only* for this condition. A more likely explanation is that thresholds in the presence of the BB modulated masker were influenced by suppression at modulation depths greater than about 0.75. Psychophysical suppression can be observed in a forward-masking paradigm when the bandwidth of the noise masker is increased beyond a critical band (e.g., Houtgast, 1974; Weber, 1978). The effect of suppression in the present study would be to lower the thresholds in the presence of the BB modulated masker, at least at large modulation depths where thresholds may be governed by forward masking. The general lack of a similar effect at 250 Hz could be due to a smaller amount of suppression at the lower frequency, a possibility we address in the General Discussion.

VI. EXPERIMENT 4: CONTRALATERAL FLANKING ENERGY

A. Rationale and conditions

As noted previously, the effect of signal frequency with the ERB masker is largely the result of differences in the overall masker level as a function of frequency, as opposed to differences in auditory processing across frequency. The same cannot be said for the BB masker, suggesting that there is an across-channel (auditory) component that is influenced by signal frequency. The results from experiment 2 (Fig. 3) indicate that the effect of signal frequency on the MUD exists primarily at modulation depths greater than about 0.75. The results from experiment 3 (Fig. 5) strongly suggest that it is only at those larger modulation depths where thresholds in the presence of a modulated masker are influenced significantly by forward masking. Taken together, these results suggest that the signal frequency effect is largely influenced

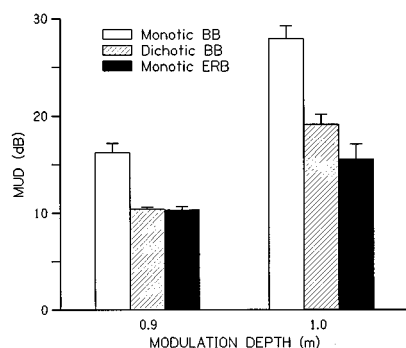


FIG. 6. MUD values obtained in three masking conditions (identified by unique histogram shading) at two different modulation depths. Error bars represent plus one standard error.

by processing revealed in forward masking. It is somewhat uncertain as to what mechanisms underlie forward masking, and whether they are peripheral (e.g., Harris and Dallos, 1979) or central in origin (e.g., Carlyon, 1988; Relkin and Turner, 1988). The purpose of the present experiment was simply to examine whether the influence of the across-channel component could be observed under a dichotic condition.

The signal was a 4000-Hz sinusoid with a duration of 500 ms. The masker was a band of noise with a duration of 500 ms. The masker was presented at a spectrum level of 40 dB SPL. It was unmodulated or SAM at a rate of 8 Hz and a depth of 0.9 or 1.0. Three conditions were run. In one (monotic ERB), the masker was one ERB wide and presented to the ear with the signal. In another (monotic BB), a notched noise was added to the ERB masker to produce a BB masker that was presented to the ear with the signal. Note that these first two conditions replicate some of the ERB and BB conditions from experiments 1 and 2, although the method of generating the BB masker was somewhat different here.³ (The thresholds in these conditions were similar to those in experiments 1 and 2.) In the third condition (dichotic BB), the notched noise was presented to the ear contralateral to that receiving the ERB masker and signal.

B. Results and discussion

The results were similar across the subjects, and thus only their mean data are presented in Fig. 6. The histograms on the left give the MUD values when the modulated masker was modulated at a depth of 0.9, whereas those on the right give the values when the depth was 1.0. As observed previously, the MUD is smaller at a depth of 0.9 than 1.0. Of interest here is the finding that the MUD value for the dichotic BB condition was similar to (within 3 dB of) that for the monotic ERB condition, indicating that the MUD increases with increasing masker bandwidth only when the masker energy flanking the ERB masker is located in the ear receiving the ERB masker and signal. This suggests that the effect of the flanking energy is mediated at a relatively peripheral site.

VII. GENERAL DISCUSSION

The major aim of the present study was to evaluate in detail the signal frequency effect that had been observed previously in the MUD paradigm (Bacon *et al.*, 1997) and to gain further insight into the possible processing underlying the frequency effect. The results from experiments 2 and 3 suggest that the signal frequency effect exists as a result of processes revealed only in forward masking. Furthermore, because the signal frequency effect is considerably larger with a BB masker than with an ERB masker (e.g., Figs. 1 and 3), it is necessary to consider explanations to account for both a within-channel and an across-channel effect of signal frequency.

A. Within-channel effect of signal frequency

As noted previously, most of the effect of signal frequency with the ERB masker was due to the effect of frequency on the threshold in the presence of the *unmodulated* masker (see Fig. 1). This effect can be understood by considering the fact that the overall level of the ERB masker increased with increasing signal frequency. (Recall that the spectrum level of the noise masker was held constant, while the bandwidth of the masker increased with increasing signal frequency.) Thus it is largely unnecessary to consider a change in within-channel processing across frequency. This conclusion is consistent with the results of our previous study (Bacon *et al.*, 1997), as well as with results of an ongoing study in which we measure the MUD as a function of frequency for a fixed-bandwidth (100-Hz) masker.

The above explanation for the effect of signal frequency on the MUD_{ERB} may seem inconsistent with the finding that the effect of signal frequency largely disappears when the modulation depth of the masker is reduced to a value below about 0.75 (Fig. 3), as it implicates processing in the presence of the *unmodulated* masker as being responsible for the within-channel signal frequency effect. It is not inconsistent, however, when one considers the finding that at modulation depths below about 0.75, thresholds are governed largely by simultaneous masking in the dips (Fig. 5). Thus at those lower depths, thresholds in the presence of the unmodulated masker and those in the presence of the modulated masker will increase with increasing signal frequency, due to the fact that *both* are similarly affected by the concomitant increase in overall masker level.

Although the effect of signal frequency with the ERB masker was largely due to the thresholds in the presence of the unmodulated masker, it is worth noting that there was a tendency for the thresholds in the presence of the *modulated* ERB masker to decrease slightly as signal frequency increased from 250 to about 1000 or 2000 Hz (see Fig. 1), suggesting that a small component of the within-channel frequency effect may be due to processing in the presence of the modulated masker. In other words, at least a small part of the frequency effect may be due to differences in auditory processing across frequency. It is important to consider what type of processing might contribute to this component of the within-channel effect of signal frequency. Given that thresholds in the presence of the modulated masker (at least at the larger depths where a frequency effect exists) are likely to be

influenced by processes revealed in forward masking, it is reasonable to consider an explanation based on forward masking. The explanation considered here is based on differences in the “transition” between simultaneous and forward masking. In particular, we suggest that some of the within-channel component may reflect smaller differences between simultaneous- and forward-masked thresholds at low than at high signal frequencies, as has been observed previously (e.g., Duifhuis, 1973; Carlyon, 1988; Moore and Glasberg, 1987), and offer two explanations for why this may be the case. A smaller difference would be reflected here as a smaller MUD_{ERB} , inasmuch as the threshold in the presence of the modulated masker is governed by forward masking.

One reason that the difference between simultaneous- and forward-masked thresholds (and hence the MUD_{ERB}) might be smaller at low rather than at high signal frequencies is related to the relatively long “ringing” of the narrow peripheral filters at low frequencies (Duifhuis, 1973; Carlyon, 1988). However, it would seem very unlikely that ringing in the auditory filters could influence the thresholds obtained here in the presence of the SAM masker modulated at a large modulation depth, given the relatively long “fall time” of the 8-Hz SAM masker. Moreover, our simulations of the output of auditory filters in response to a masker with such a long rise–fall time suggest that the ringing is minimal and would almost certainly not influence threshold.

Another, more likely, reason that the MUD_{ERB} would be smaller at low signal frequencies is related to the difference in the sensation level (SL) of the ERB maskers. The rate of recovery in forward masking depends upon the SL of the masker (Plomp, 1964; Widin and Viemeister, 1979; Jesteadt *et al.*, 1982; Moore and Glasberg, 1983; Nelson and Freyman, 1987), being slower at lower SLs. Although we did not measure the quiet threshold for the ERB maskers, it is likely that the SL for the ERB masker centered at 250 Hz was about 25 dB less than that for the ERB masker centered at 4000 Hz. This is based on the normal audibility curve (the threshold for a 4000-Hz tone is about 15 dB lower than that for a 250-Hz tone) and the fact that the overall level of the ERB masker was about 9 dB greater at 4000 Hz than at 250 Hz. Consistent with this explanation are some results from our previous study (Bacon *et al.*, 1997; Fig. 9), which indicate that a 25-dB decrease in the level of an ERB masker centered at 4000 Hz would reduce the MUD_{ERB} by about 5 or so dB, which would tend to equate the size of the MUD_{ERB} at 250 and 4000 Hz.

B. Across-channel effect of signal frequency

The extent of the across-channel processing is revealed in the MUD paradigm by comparing the MUD_{BB} with the MUD_{ERB} . The difference between these two is a measure of CMR. The intent here is not to review different models or theories for CMR (see Green, 1993, for such a review), although it is likely that the processing revealed in other studies of CMR is also involved in the present study. Instead, the focus here is on the mechanism responsible for the across-channel effect of signal frequency. We suggested previously (Bacon *et al.* 1997; also Sec. V B) that suppression might be

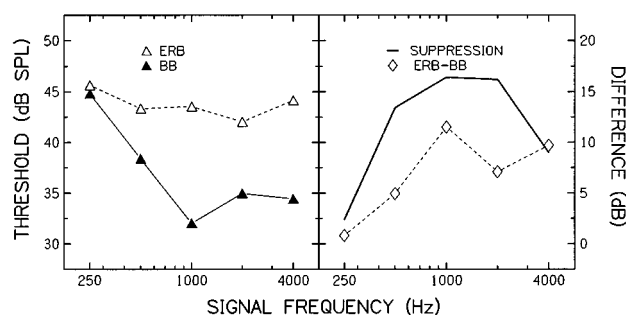


FIG. 7. In the left panel, thresholds in the presence of a masker modulated at a rate of 8 Hz and a depth of 1.0 are plotted as a function of signal frequency. The difference between those thresholds is plotted (diamonds) in the right panel. These results (from experiment 1) have been averaged across two of the three subjects. In addition, the solid line without symbols in the right panel represents the amount of psychophysical suppression measured in the same two subjects (Lee and Bacon, in preparation). See text for more details.

responsible for the across-channel effect of signal frequency. We consider that possibility in more detail here.

The following findings are consistent with the possibility that the across-channel effect of frequency may be related to differences in psychophysical suppression across frequency: (1) the largest across-channel effects of signal frequency are observed at the largest modulation depths (Fig. 4), where forward masking apparently plays a significant role; (2) the amount of CMR (at least at 1000 and 4000 Hz, where there is a clear CMR) increases significantly as the modulation depth of the masker increases beyond 0.75 (Fig. 4), suggesting the contribution of an additional across-channel mechanism at a depth where forward masking plays a role; (3) the across-channel effect in the present study exists only when the masker and flanking energy are presented to the same ear (Fig. 6); and (4) the results from other studies (Shannon, 1976; Thibodeau and Fagelson, 1993; Lee and Bacon, in preparation) suggest that psychophysical suppression may indeed be larger at high than at low signal frequencies.

To illustrate how both the across-channel processing revealed in the present experiments and psychophysical suppression are affected by signal frequency, we have plotted results from the present study along with results from a suppression study (Lee and Bacon, in preparation) in Fig. 7. The results are averaged across the two subjects who participated in both studies. In the left panel (referred to the left-hand axis), thresholds from the present study (experiment 1, modulation rate of 8 Hz) are plotted as a function of signal frequency for both the ERB and BB masker (unfilled and filled triangles, respectively). The maskers were modulated at a depth of 1.0, and thus these represent thresholds that are thought to be largely influenced by forward masking. In the right panel (referred to the right-hand axis), the unfilled diamonds represent the difference between the thresholds plotted in the left panel. This difference, which can be thought to represent the influence of an across-channel process, increases from about 0 dB at 250 Hz to about 10 dB at frequencies from 1000 to 4000 Hz. Also plotted in the right panel as a solid line without symbols are our estimates of psychophysical suppression obtained in a forward-masking

paradigm.⁴ Although the two sets of data do not generally overlap, which is probably not surprising given the numerous differences between the two paradigms, they do display similar effects of signal frequency. This is at least consistent with the possibility that the across-channel effect of signal frequency observed in the MUD paradigm reflects differences in suppression across frequency.

One difference between the two functions in the right panel of Fig. 7 that may be important is that the magnitude of suppression reaches or nearly reaches an asymptote at a signal frequency of 500 Hz, whereas the difference between the thresholds in the MUD paradigm increases up to a frequency of 1000 Hz. Whether this simply reflects, for example, stimulus differences between the two paradigms, or whether it argues against suppression as an explanation for the across-channel effect of signal frequency observed in the MUD paradigm is unclear. If suppression is not involved, then some other mechanism that is also revealed in forward masking (or at least only at large modulation depths) must account for the across-channel effect of frequency. At present, we are unaware of any other likely candidates. Further experiments should help clarify the role that suppression may play in MUD experiments with relatively BB maskers.

ACKNOWLEDGMENTS

This research was supported in part by a grant from NIDCD (DC01376). We thank Wes Grantham and two anonymous reviewers for their helpful comments on a previous version of this manuscript.

¹The maskers were modulated after bandlimiting, and thus their bandwidths were somewhat wider when they were modulated than when they were unmodulated. Because the modulation rate was typically 8 Hz or less, the differences in bandwidth were relatively small, and probably did not influence significantly the results presented here.

²Although the 30-ms signal used in the present experiment *may* be similar to the effective signal duration during a *single* dip of the modulated masker in experiment 2, there are four such dips in an observation interval, and thus the combined effective duration would be longer than 30 ms, perhaps as long as 120 ms. Regardless, if we assume that the effective signal duration, whatever it may be, is more or less unaffected by changes in modulation depth from 0.25 to 1.0, then it is possible to make meaningful comparisons across experiments (and hence signal durations) by examining changes in the relative thresholds.

³The 460-Hz-wide ERB masker was generated as before. The BB masker was generated as follows: First, a 10-kHz-wide masker was generated (the BB masker in the previous experiments), and then a notch was produced in that noise by low-pass (3.5 kHz) and high-pass (4.65 kHz) filtering (Kemo VBF 25.01; 135 dB/oct) the masker in parallel, and then adding the outputs of those filters together. Finally, this notched masker was added to the ERB masker to create a BB masker. The low-pass and high-pass cutoff frequencies were chosen so that the addition of the notched noise would have little or no effect on the ERB masker and so that the spectrum of the resultant BB masker would be relatively flat.

⁴Suppression was estimated as the difference between the threshold obtained with a BB masker and that obtained with an ERB masker. The onset of the 20-ms signal followed the offset of the 500-ms masker by 20 ms to minimize confusion effects that can exist in forward masking with narrow-band noise maskers (e.g., Moore and Glasberg, 1985; Neff, 1985).

ANSI (1989). ANSI S3.6-1989, "Specifications for audiometers" (American National Standards Institute, New York).

Bacon, S. P., and Lee, J. (1995). "Temporal resolution and CMR can depend upon frequency," *J. Acoust. Soc. Am.* **97**, 3273.

- Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (1997). "Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level," *J. Acoust. Soc. Am.* **101**, 1600–1610.
- Carlyon, R. P. (1988). "The development and decline of forward masking," *Hearing Res.* **32**, 65–80.
- Carlyon, R. P., Buus, S., and Florentine, M. (1989). "Comodulation masking release for three types of modulator as a function of modulation rate," *Hearing Res.* **42**, 37–46.
- Duifhuis, H. (1973). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," *J. Acoust. Soc. Am.* **54**, 1471–1488.
- Eddins, D. A., and Wright, B. A. (1994). "Comodulation masking release for single and multiple rates of envelope fluctuation," *J. Acoust. Soc. Am.* **96**, 3432–3442.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.* **47**, 103–138.
- Green, D. M. (1993). "Auditory intensity discrimination," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, Berlin).
- Grose, J. H., and Hall, III, J. W. (1989). "Comodulation masking release using SAM tonal complex maskers: Effects of modulation depth and signal position," *J. Acoust. Soc. Am.* **85**, 1276–1284.
- Haggard, M. P., Hall, III, J. W., and Grose, J. H. (1990). "Comodulation masking release as a function of bandwidth and test frequency," *J. Acoust. Soc. Am.* **88**, 113–118.
- Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **76**, 50–56.
- Harris, D. M., and Dallos, P. (1979). "Forward masking of auditory nerve fiber responses," *J. Neurophysiol.* **42**, 1083–1107.
- Hawkins, Jr., J. E., and Stevens, S. S. (1950). "The masking of pure tones and of speech by white noise," *J. Acoust. Soc. Am.* **22**, 6–13.
- Hicks, M. L., and Bacon, S. P. (1995). "Some factors influencing comodulation masking release and across-channel masking," *J. Acoust. Soc. Am.* **98**, 2504–2514.
- Houtgast, T. (1974). "Lateral suppression in hearing," Doctoral thesis, Academische Pers B. V.
- Jesteadt, W., Bacon, S. P., and Lehman, J. R. (1982). "Forward masking as a function of frequency, masker level, and signal delay," *J. Acoust. Soc. Am.* **71**, 950–962.
- Lee, J., and Bacon, S. P. (in preparation). "Measures of psychophysical suppression as a function of frequency," *J. Acoust. Soc. Am.*
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Growth of forward masking for sinusoidal and noise maskers as a function of signal delay; implications for suppression in noise," *J. Acoust. Soc. Am.* **73**, 1249–1259.
- Moore, B. C. J., and Glasberg, B. R. (1985). "The danger of using narrow-band noise maskers to measure 'suppression'," *J. Acoust. Soc. Am.* **77**, 2137–2141.
- Moore, B. C. J., and Glasberg, B. R. (1987). "Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes," *J. Acoust. Soc. Am.* **82**, 69–79.
- Neff, D. L. (1985). "Stimulus parameters governing confusion effects in forward masking," *J. Acoust. Soc. Am.* **78**, 1966–1976.
- Nelson, D. A., and Freyman, R. L. (1987). "Temporal resolution in sensorineural hearing-impaired listeners," *J. Acoust. Soc. Am.* **81**, 709–720.
- Plomp, R. (1964). "Rate of decay of auditory sensation," *J. Acoust. Soc. Am.* **36**, 277–282.
- Relkin, E. M., and Turner, C. W. (1988). "A reexamination of forward masking in the auditory nerve," *J. Acoust. Soc. Am.* **84**, 584–591.
- Schooneveldt, G. P., and Moore, B. C. J. (1987). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," *J. Acoust. Soc. Am.* **82**, 1944–1956.
- Shannon, R. V. (1976). "Two-tone unmasking and suppression in a forward-masking situation," *J. Acoust. Soc. Am.* **59**, 1460–1470.
- Thibodeau, L. M., and Fagelson, M. A. (1993). "Comparison of suppression across frequencies," *J. Acoust. Soc. Am.* **93**, 2314(A).
- Weber, D. L. (1978). "Suppression and critical bands in band-limiting experiments," *J. Acoust. Soc. Am.* **64**, 141–149.
- Widin, G. P., and Viemeister, N. F. (1979). "Intensive and temporal effects in pure-tone forward masking," *J. Acoust. Soc. Am.* **66**, 388–395.

Sensitivity to changes in overall level and spectral shape: An evaluation of a channel model

Jennifer J. Lentz

Department of Bioengineering, 120 Hayden Hall, University of Pennsylvania, Philadelphia,
Pennsylvania 19104

Virginia M. Richards

Department of Psychology, 3815 Walnut Street, University of Pennsylvania, Philadelphia,
Pennsylvania 19104

(Received 28 June 1996; revised 5 February 1997; accepted 14 February 1997)

Two experiments involving level and spectral shape discrimination which test an optimal channel model developed by Durlach *et al.* [J. Acoust. Soc. Am. **80**, 63–72 (1986)] are described. The model specifies how the auditory system compares and/or combines intensity information in different frequency channels. In the first experiment, psychometric functions were obtained for the discrimination of changes in level and discrimination of changes in spectral shape for an eight-tone complex sound. A variety of different base spectral shapes were tested. In some conditions, level randomization was introduced to reduce the reliability of across-interval changes in level. Increasing the amount of level variation degraded performance for the level discrimination task but had no effect on the shape discrimination task. In all conditions, sensitivity to changes in spectral shape was superior to sensitivity to changes in level. Consequently, two models of central noise are evaluated in an attempt to explain these results; one in which central noise acts prior to the formation of the likelihood ratio and one in which central noise degrades the likelihood ratio. The former model is more successful in accounting for the data. In a second experiment, the detectability of a level increment to one component of a multitone complex was measured. The frequency content of the complex was varied by systematically removing six components from a 23-component complex. Thresholds were measured for increments at three different signal frequencies. A common trend in the data was that when there was a spectral gap directly above the signal frequency, thresholds were lowest. This result differs from the predictions of a simple channel model, and contrasts with results presented by Green and Berg [Q. J. Exp. Psychol. **43A**, 449–458 (1991)]. © 1997 Acoustical Society of America. [S0001-4966(97)01806-7]

PACS numbers: 43.66.Ba, 43.66.Fe, 43.66.Jh [WJ]

INTRODUCTION

Profile analysis experiments examine the ability to discriminate a change in the spectral shape of a multitonal complex (Green, 1988b). In a common profile analysis experiment, observers might discriminate between a standard stimulus that is comprised of many equal-amplitude tones and a signal stimulus in which a tone has been added in-phase to one of the components of the standard. For profile analysis tasks, observers could potentially make a detection decision based on a single component. To reduce the reliability of the single-component cues, experimenters typically choose the level of each stimulus at random prior to each presentation. As a result of this level randomization, changes in either overall level or the level of a single component may be discounted as contributing to detection, insuring that the overall spectral shape of the stimuli forms the basis of detection.

One model available to account for performance in spectral shape discrimination experiments is the channel model of Durlach *et al.* (1986). Durlach *et al.* describe a maximum-likelihood estimator that provides predictions for changes in overall level and/or changes in the spectral shape of complex sounds. The basic model assumes the level/shape decision proceeds by (a) first resolving the input into frequency bands

via a linear filter bank, (b) computing the level in each frequency band, (c) introducing “channel noise” by adding independent Gaussian noise to each level estimate, and (d) computing a decision variable via log likelihood estimators. As will be described later, Durlach *et al.* provide two means of incorporating central noise: one in which deviates are added prior to the formation of the likelihood ratio and one in which a deviate is added after the formation of the likelihood ratio.

Even though Durlach *et al.*’s (1986) model has relatively few parameters and is explicit in providing predictions for a variety of experiments, relatively few experimental tests of the model are available. In an experiment in which observers discriminated between the sum of several equal-amplitude tones and a tophat-shaped stimulus (the middle frequency components were incremented above a mean level; the outer frequency components decremented in level), Green (1992) systematically increased the number of components while keeping the frequency range constant. Thresholds fell as the square root of the number of components as predicted by the channel model. Another test of the channel model was completed by Berg and Green (1990) who evaluated the detectability of a tone added to the middle component of one of three standards, consisting of the sum of either

3, 5, or 11 equal amplitude tones. They found that the relative weight (or contribution) of each of the components depended on the total number of components in the way predicted by the channel model. In addition to the studies described above, Farrar *et al.* (1987) applied the channel model to broadband stimuli, specifically speechlike sounds (unvoiced plosives and fricatives). The channel model also provided generally good fits to their data.

The current experiments were completed to provide a more extensive test of Durlach *et al.*'s (1986) channel model. In the first experiment, psychometric functions were obtained for both level and spectral shape discrimination tasks. All stimuli were comprised of eight tones. For the level discrimination task, all eight components of the complex were incremented by Δ dB. For the shape discrimination task, sensitivity to changes in spectral shape was measured (e.g., profile analysis). The shape task employed "balanced" increments and decrements in the constituent components of the eight-tone complex. Durlach *et al.* defined a balanced stimulus as one satisfying the equation $\sum_i (\Delta_i / \sigma_i) = 0$, with i denoting the channel number, Δ_i the increment/decrement in level in dB, and σ_i^2 the variance associated with channel (or encoding) noise. Here, it is assumed that the σ_i are the same for all i , and so the definition of the balanced stimulus becomes $\sum_i \Delta_i = 0$. An additional constraint placed on the shape discrimination stimuli was that $\Delta_i = \pm \Delta$, with the number of increments equaling the number of decrements. Thus, in both the level and shape discrimination tasks the signal stimulus differed from the standard stimulus in that there was a Δ dB change in the level of each component. As a result of the consistent change in level on a component by component basis, Durlach *et al.*'s model predicts that d'_L and d'_S , the d' values for the level and shape tasks, respectively, will be equal. The details of this argument will be discussed following the initial data presentation.

For the particular stimulus construction chosen for the experiment, when level variation was not present, sensitivity to changes in spectral shape was expected to be equal to sensitivity to changes in overall level. This prediction rests on an assumption that the parameters specified in the channel model are the same for both level and shape discrimination. The model also provides predictions concerning the magnitude of the effect of interval-by-interval level randomization on level discrimination and also predicts no effect of level randomization in the shape discrimination task. The latter prediction has been empirically demonstrated (Farrar *et al.*, 1987; Green, 1988b; Mason *et al.*, 1984).

In a second experiment, the threshold for the detection of an increment in level of a single component of an equal-amplitude standard was measured. Several signal frequencies were tested, and the standard was a multitone complex with different spacings between component tones. For the channel model, if the σ_i are equal for all i , if processing for all channels is the same, and if all pairs of channels share the same covariances (including mutual independence), then the detection of an increment is expected to be independent of which component bears the signal and depends on the num-

ber of spectral components, not their distribution in frequency. Both predictions were tested in this experiment.

Past work shows that the former prediction, that the detection of an increment should be independent of the component containing the increment, is incorrect. For example, when standards with components ranging in frequency from 200 to 5000 Hz in logarithmic steps are used, maximal sensitivity occurs for midfrequency signals (near 1000 Hz, the geometric center of the profile; Green and Mason, 1985; Bernstein and Green, 1987; Green, 1988a; Green *et al.*, 1987). Because thresholds are lowest for midfrequency signals and increase for signal frequencies more disparate from the center, this pattern of results is sometimes referred to as the profile bowl. For the channel model, this result might be taken to imply that the σ_i are not equal for all i .

Using the COSS (Conditional On a Single Stimulus) analysis technique, Green and Berg (1991) expanded on the results of Berg and Green (1990) to include detection of a signal at various frequencies. After estimating the relative weights of the different frequency regions, they compared their results against the predictions of the channel model (Durlach *et al.*, 1986). They concluded that the observers behaved as predicted by the channel model for the central 1000-Hz signal frequency. In contrast, for noncentral signals, suboptimal weighting schemes appear to have been adopted by the observers, but the data showed no noticeable consistency in the strategies adopted by the individual observers.

In contrast to the number of experiments that have evaluated the effects of signal frequency on sensitivity to changes in spectral shape, relatively little work has been completed examining the role that the distribution in frequency of the spectral components has on sensitivity. Green (1988a) reported that thresholds for the detection of an increment to one component of a three-component standard were relatively independent of the frequency spacing of the constituent components. A primary goal of experiment II was to provide a more complete evaluation of the role that spectral content of the standard plays in the detection of an amplitude increment of a single component.

For experiment II, the standard profile consisted of either the sum of 23 or 17 equal-amplitude tones. The 17-component complex was generated by systematically excluding six components of the 23-component complex. Thus, spectral "gaps" in the profile were introduced. By evaluating the relationship between the "gap" frequency and the threshold for the detection of an increment in level of a single component, we hoped to obtain a rough estimate of the relative importance of different frequency regions. For the 17-component complex, two distinct types of profiles were tested: (a) those with gaps placed symmetrically with respect to the signal frequency, and (b) those with gaps placed either above or below the signal frequency. If the predictions made by the channel model hold and assuming all σ_i are equal, thresholds would not depend on the placement of the gap.

I. EXPERIMENT I: SENSITIVITY TO CHANGES IN LEVEL VERSUS CHANGES IN SPECTRAL SHAPE

A. Methods for experiment I

1. Stimuli

The stimuli consisted of the sum of eight tones, ranging in frequency from 300 to 3000 Hz, with equidistant spacing on a logarithmic scale. The mean level of each tone was 55 dB SPL. On each presentation, the phase of each tone was randomly selected from a uniform distribution ranging from 0 to 2π rad. Observers were tested on two types of tasks. One task consisted of detecting a change in the overall level of the stimulus, while the second task consisted of detecting a change in spectral shape. These will be considered in turn.

a. Level task. Observers detected an overall increment in level (each component of the eight-tone complex was altered by the same amount) for three different types of stimuli; that is, for complexes with three different spectral shapes. If the spectral shape of the standard matters, our instantiation of the channel model is incorrect and cannot be reasonably applied to the remaining portion of the experiment. The three different spectral shapes tested were (1) *flat*, all tones were equal in amplitude; (2) *tophat*, the four tones in the center of the frequency range of the complex were 1 dB above the mean level, and the two low-frequency and two high-frequency tones were 1 dB below the mean level; (3) *inverted tophat*, opposite to the *tophat* condition, the four tones in the center of the complex were 1 dB below the mean level, and the two low- and two high-frequency tones were 1 dB above the mean level.

The standard was produced by decrementing all tones by $\Delta/2$ dB, and the signal was produced by incrementing all tones by $\Delta/2$ dB. The level change Δ is the difference in decibels of a single component of the signal stimulus with respect to the corresponding component of the standard stimulus, as defined by Durlach *et al.* (1986). Note that both the signal and standard had the same spectral shape.

b. Shape task. Observers were asked to detect changes in spectral shape for two different stimulus configurations. These two configurations were both “balanced” stimuli, as described by Durlach *et al.* (1986). The two shape discriminations were (1) *tophat* versus *flat* and (2) *tophat* versus *inverted tophat*. (1) For *tophat* versus *flat*, the standard stimulus (*flat*) consisted of equal-amplitude tones. For the signal stimulus (*tophat*), the four components at the four central frequencies were increased by Δ dB, and the two lowest-frequency and two highest-frequency tones decremented by Δ dB. The resultant change in level for each spectral component was Δ dB. (2) For *tophat* versus *inverted tophat*, the standard (*inverted tophat*) had the middle and edge components decremented and incremented by $\Delta/2$ dB, respectively. For the signal (*tophat*), the middle and edge components were incremented and decremented by $\Delta/2$ dB, respectively. Again, the level of each component changed by Δ dB.

c. Level randomization. In the level and shape tasks, psychometric functions were obtained both with and without across-interval level randomization. On each stimulus presentation, a random perturbation of the overall level was chosen from a normal distribution with a mean of 0 and a

standard deviation of either 0 (no rove), 0.3, or 0.7 dB. These will be referred to as the 0-, 0.3-, and 0.7-rove conditions. Due to a minor roundoff error in the program, these distributions were just marginally different from being normally distributed. It should be noted that the standard deviations selected for the rove distributions (0–0.7 dB) are small in comparison to the range of level variation commonly used in shape discrimination tasks (20–40 dB, Green, 1988b). In order to test the predictions of the model, it was necessary to keep the values of Δ constant for both the shape and level tasks. Due to the detrimental affects of small amounts of rove on the level task, small standard deviations were required.

2. Procedure

The stimuli were generated digitally, played by a two-channel digital-to-analog converter (DAC, TDT DA1) at a sampling rate of 20 kHz, and low-pass filtered at 5 kHz (TDT FT5; approximately 125 dB/oct falloff). The total duration of each stimulus was 100 ms, including 10-ms cosine squared rise/decay ramps. The interstimulus interval duration was approximately 750 ms.

Stimuli were presented diotically through Sennheiser HD 410 SL earphones to observers seated in a single-walled, sound-attenuated booth. A three-alternative, forced-choice procedure was used because pilot work uncovered confusions as to which stimulus was the signal and which stimulus was the standard. The signal was as likely to occur in any one of the three intervals. Percent correct was calculated for 50 trials with Δ as the independent variable. d' values were computed for each 50-trial estimate (Macmillan and Creelman, 1991). The values reported are the average of four estimates. Four to six different Δ 's were tested to produce psychometric functions.

3. Observers and order of data collection

Three observers, ranging in age from 24 to 28, were tested. All observers had pure-tone hearing thresholds better than 15 dB HL in the range of 250–8000 Hz. Observer 3, the first author, had prior experience in psychoacoustic experiments including profilelike tasks. The other two observers were naive to psychoacoustic experiments. After completing at least ten hours of practice in both the shape and level tasks, data collection began. For the 0-rove condition, observers first completed the level—*flat* condition (i.e., the level discrimination task with a flat spectral shape), and second completed the shape—*tophat* versus *flat* condition (i.e., a shape discrimination task). The level—*flat* condition was tested first so as to select suprathreshold values of Δ for the remaining conditions. The shape—*tophat* versus *flat* condition was tested second as a means of checking that the *tophat* profile was indeed perceptually distinct from *flat*. Observers completed the 0-rove conditions first, the 0.7-rove conditions second, and the 0.3-rove conditions last. For the 0.3- and 0.7-rove conditions, psychometric functions were completed for one condition before moving onto the next, and the order in which the conditions were tested was random for each observer. Finally, because there was little effect of spectral

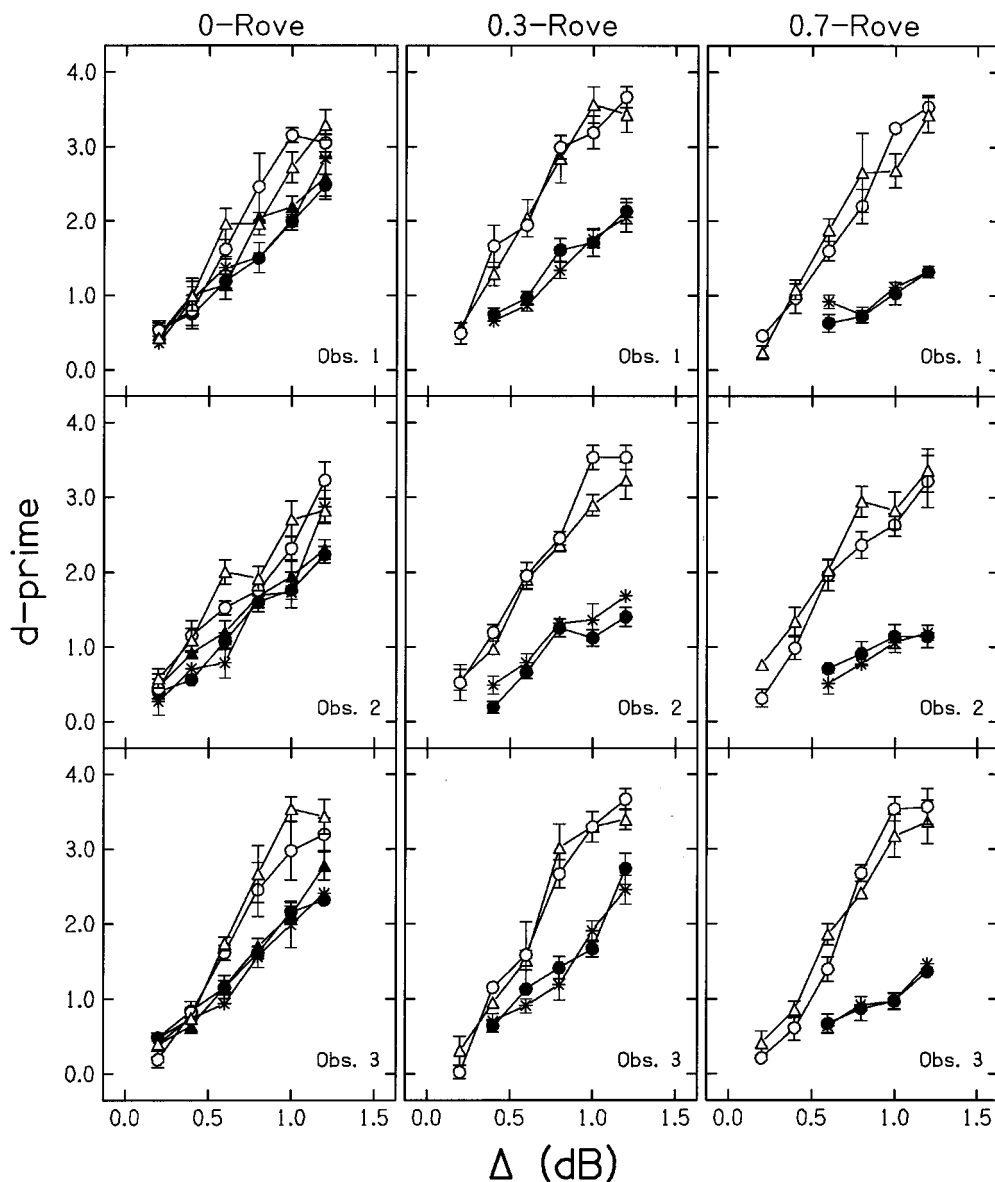


FIG. 1. d' is plotted as a function of the level increment Δ in dB. Panels in different rows are for individual observers. Panels in different columns are for different degrees of interval-by-interval level randomization: The standard deviation of the rove was 0 dB (no rove, left column), 0.3 dB (middle), or 0.7 dB (right). Filled symbols indicate d' scores for the level task, and open symbols indicate d' scores for the shape task. For the level task, the stimulus shapes are *flat* (asterisks), *tophat* (filled circles), and *inverted tophat* (filled triangles). For the shape task, the discrimination is *tophat* versus *flat* (open circles) or *tophat* versus *inverted tophat* (open triangles). Error bars indicate the standard errors of the mean across four d' estimates.

shape for the level task, sensitivity to changes in overall level of the *inverted tophat* stimulus was completed only for the 0-rove condition.

B. Results and discussion of experiment I

1. Psychophysical results

Figure 1 shows the psychometric functions measured for the three observers (rows) and for three different amounts of level variation (columns). d' is plotted as a function of Δ .¹ Error bars indicate standard errors of the mean across four replicate estimates of d' . Open symbols show the results in the shape conditions and filled symbols the level conditions. For the level conditions, the asterisks, filled circles, and filled triangles represent the stimulus shapes of *flat*, *tophat*, and *inverted tophat*, respectively. For the shape functions, open

circles represent *tophat* versus *flat*, and the open triangles represent *tophat* versus *inverted tophat* conditions. The left panel shows psychometric functions obtained without level variation (0-rove). The middle panel shows functions for the 0.3-rove condition, and the right panel shows functions for the 0.7-rove condition.

As the degree of level randomization grows, d' for a given Δ falls for the level discrimination task but is unaltered in the shape discrimination task. In both the level and shape discrimination tasks, the psychometric functions are invariant with respect to stimulus shape. That is, as predicted by the channel model, no effect of spectral shape is apparent when balanced stimuli are tested. Sensitivity to changes in shape is better than the sensitivity to changes in level. When no interval-by-interval level randomization is used, the chan-

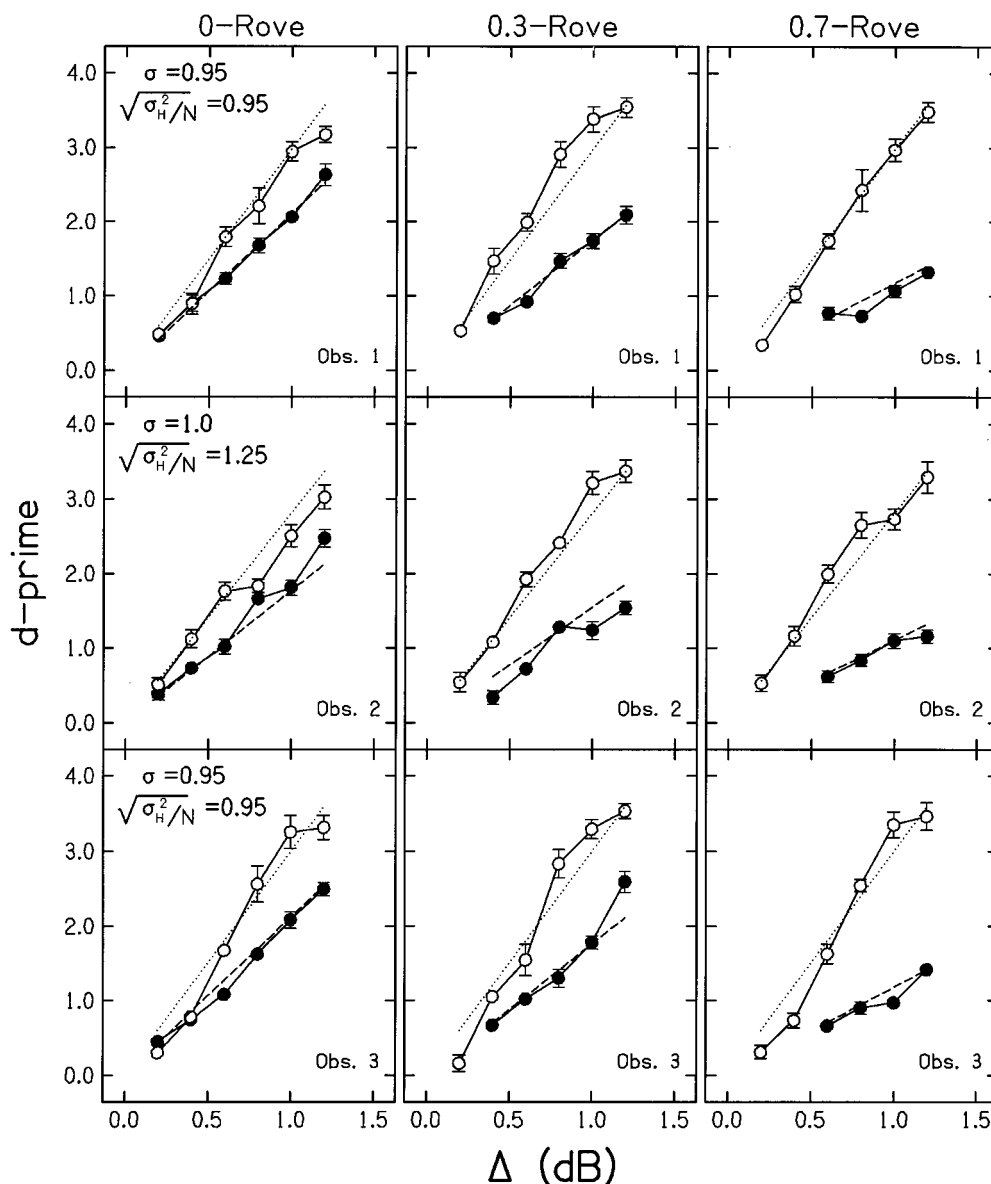


FIG. 2. Average psychometric functions for the three observers (rows) and the three degrees of level randomization (0 dB, left; 0.3 dB, center; and 0.7 dB, right) are presented by plotting d' as a function of level change, Δ in dB. Open circles are for the shape discrimination task and closed circles for the level discrimination task. The dotted and dashed lines are the best-fitting predictions as generated by Durlach *et al.*'s (1986) channel model for the shape and level discrimination tasks, respectively. Error bars indicate standard errors of the mean across d' estimates. Values obtained for σ (encoding noise) and $\sqrt{\sigma_H^2/N}$ (central noise) are indicated in the upper left corners of the leftmost panels for each observer.

nel model predicts a different result: Equal values of Δ should lead to equal d' values if all parameters defined by the model are invariant with respect to either level or shape discrimination.

One possible explanation of the difference in the psychometric functions for the level and shape tasks at 0-rove is suggested by the results reported by Berliner and Durlach (1972) and Green *et al.* (1983). Sensitivity to differences in level depends on the duration of the interstimulus interval, but sensitivity to differences in spectral shape is relatively free of such effects. Presumably this difference reflects differences in memory load, possibly in conjunction with differences in the representations employed in the shape and level tasks (Green *et al.*, 1983; Kidd *et al.*, 1988). The paucity of different spectral shapes tested in the current experi-

ment may have reduced the observer's dependence on across-interval comparisons in the spectral shape task (Kidd *et al.*, 1988). At present, the reasons for the differences in sensitivity in the shape and level tasks (0-rove) are not clear; in our evaluation of the channel model, we simply treat the difference as reflecting differences in central noise rather than modeling particular aspects of perceptual processing.

2. Evaluation of several forms of the channel model

Figure 2 shows average psychometric functions. For each observer, d' scores are collapsed across the different shape configurations, resulting in average shape (open circles) and level (filled circles) functions. The dotted and

dashed lines show the predictions of one instantiation of Durlach *et al.*'s (1986) channel model, which is briefly developed below.

In a simple version of the channel model, the stimulus components are assumed to activate independent channels, and the representation of level in each channel is corrupted by a Gaussian encoding noise of mean zero and variance σ^2 (the only free parameter in this simple version of the model). This corruption will be described as reflecting channel noise. The corrupted level estimates are then used to form a likelihood ratio, and an ideal decision is made. Equation (6) of Durlach *et al.* (1986) shows the index of sensitivity d' as

$$d' = \frac{1}{\sigma} \left[\sum_{i=1}^N \Delta_i^2 - \frac{\sigma_R^2}{\sigma^2 + N\sigma_R^2} \left(\sum_{i=1}^N \Delta_i \right)^2 \right]^{1/2},$$

where N is the number of channels activated, Δ_i is the mean change in level of the i th component, and σ_R^2 is the covariance between pairs of channels due to common noise, including level randomization. Taking into consideration the balanced stimulus construction, the constraint for the shape stimuli where $\Delta_i = \pm \Delta$, and the constraint on the level stimuli where $\Delta_i = \Delta$, Durlach *et al.* (1986) show that the d' values for the level and shape tasks are:

$$d'_L = \frac{\sqrt{N}\Delta}{\sqrt{\sigma^2 + N\sigma_R^2}} \quad (1)$$

and

$$d'_S = \sqrt{N} \left(\frac{\Delta}{\sigma} \right). \quad (2)$$

Equations (1) and (2) reproduce Eqs. (12) and (14) of Durlach *et al.* (1986). Here it is assumed that σ_R^2 is known and equal to the variance of the experimenter-applied Gaussian level randomization. Note that increasing σ_R^2 reduces d'_L but does not influence d'_S . One way to think about the optimal model for the shape discrimination task involves comparing the level in each channel against the average of all channels (Berg and Green, 1990; Green, 1992). As such, the shape discrimination task is expected to be independent of the magnitude of level randomization. Finally, if σ_R^2 is zero, the level and shape tasks are expected to produce equal sensitivity, that defined by Eq. (2).

There are several assumptions implicit in Eqs. (1) and (2) that deserve consideration. First, for large values of σ_R^2 , the failure of Weber's law will reduce the accuracy of Eqs. (1) and (2) (see also Berliner and Durlach, 1972). Because the current experiment uses only small ranges of level randomization, the failure of Weber's law would not seem to warrant concern. Second, the assumption of independent channels is an important one. For multitonal stimuli with components separated by frequency ratios greater than 1.2 to 1.4, Green and colleagues (e.g., Green *et al.*, 1983; Green and Mason, 1985) report an absence of single-channel masking for the detection of an increment of the 1000-Hz component of the standard. It seems reasonable to assume, then, when the ratio between neighboring components is 1.39, as in the current experiment, the channels may be treated as

independent. Finally, Eqs. (1) and (2) reflect the assumption that all aspects of processing are the same across all activated channels.

The fact that the psychometric functions measured in the 0-rove condition differ for the shape and level tasks implies that Eqs. (1) and (2) cannot account for the results [Eqs. (1) and (2) are identical when $\sigma_R^2 = 0$]. For that reason, two additional models developed by Durlach *et al.* (1986) are considered. Both allow the inclusion of a central or decision noise. In our instantiation of these models, we make the assumption that additional noise is present in the level discrimination task, but not the shape discrimination task, reflecting the relatively poorer performance in the level task.² For the level task, the central noise is included in two ways. For the first model of central noise, it is assumed that the central noise acts prior to the formation of the likelihood ratio (e.g., additional memory noise might be modeled this way; Durlach and Braida, 1969). For the second, it is assumed that the central noise occurs after the formation of the likelihood ratio (e.g., criterion noise). For these two models, and for our stimulus configurations, the predictions derived by Durlach *et al.* (1986) are:

$$d'_L = \frac{\sqrt{N}\Delta}{\sqrt{(\sigma^2 + (\sigma_H^2/N) + N\sigma_R^2)}} \quad (3)$$

and

$$d'_L = \frac{\sqrt{N}\Delta / \sqrt{\sigma^2 + N\sigma_R^2}}{\sqrt{1 + \sigma_C^2(\sigma^2 + N\sigma_R^2)/N\Delta^2}}. \quad (4)$$

Note that for these two models, different symbols (σ_H^2 and σ_C^2) are used to denote the "central noise." Equations (3) and (4) are equivalent to Durlach *et al.*'s (1986) Eq. (21) and (17), respectively.

While the result described in Eq. (4) is for the case in which a deviate is added to the decision variable, the result described in Eq. (3) was derived (Durlach *et al.*, 1986) for the case that equal-variance internal noise (with zero mean and variance σ_H^2) is added to a transformed variable.³ Because N does not vary in the current experiment, Eq. (3) is equivalent to the case that a zero-mean Gaussian deviate is added to each channel's level estimate (the added deviate having variance σ_H^2/N). By this interpretation, there is an increase in channel noise, possibly reflecting increased memory noise (Durlach and Braida, 1969). It should also be noted that this model of central noise will yield the same predictions as a model in which the added noise is correlated across channels.

The different models were fitted to the data using a least-squares criterion. For each observer, an average of all shape psychometric functions was generated. Because d'_S , N (number of components), and Δ (the level change) are known, fitting the average shape psychometric function to Eq. (2) led to an estimate of σ for each observer. The resulting best-fitting psychometric functions are displayed in Fig. 2 as dotted lines. For each observer, the best fitting σ is indicated in the upper left corners of the leftmost panels.

After σ was estimated for each observer, psychometric functions for the level discrimination task were generated

using Eqs. (3) and (4). The free parameters σ_H^2/N and σ_C^2 were estimated using a least-squares criterion applied simultaneously to psychometric functions obtained for the three different degrees of rove (σ_R^2). The dashed lines in Fig. 2 show the fits to Eq. (3). For each observer, the best fitting $\sqrt{\sigma_H^2/N}$ is indicated in the leftmost panel of the figure. The fits provided by Eq. (3) were superior to those provided by Eq. (4).

Across the three observers, the values of σ (channel noise) ranged from 0.95 to 1.0 dB. These values were somewhat smaller than the values of 1.8 to 1.9 estimated from Green's (1992) results for balanced stimuli. For the shape psychometric functions, Eq. (2) accounted for 95%–99% of the variance. The estimated values of $\sqrt{\sigma_H^2/N}$ ranged from 0.95 to 1.25 dB. The percent of variance accounted for by the predictions based on Eq. (3) ranged from 89% to 98%. For Eq. (4), which was less successful in accounting for the level data, the values of σ_C ranged from 1.68 to 2.24, accounting for only 71%–73% of the variance in the data.

As seen in Fig. 2, and in terms of the proportion of the variance accounted by the fits, Eqs. (2) and (3) capture the data well. When the models are constrained to use a single value of channel noise for both level and shape discrimination tasks, good fits are achieved only after including a central noise parameter in the level, but not the shape, predictions. Of the two models of central noise formalized by Durlach *et al.* (1986), one in which the central noise acts prior to the formation of the likelihood ratio and one in which central noise acts after the formation of the likelihood ratio, the former model is more successful in accounting for the psychophysical data. In our instantiation, the inclusion of the central noise acts to increase the variance in each channel; thus the effective channel variance in the level discrimination task is larger than in the shape discrimination task. These results are consistent with the dependence of level, but not shape, discrimination on interstimulus duration, level variation, etc. (e.g., Berliner and Durlach, 1972; Green *et al.*, 1983; Kidd *et al.*, 1988). In particular, Berliner and Durlach (1972) show that for intensity discrimination of a 1000-Hz tone, sensitivity at an interstimulus duration of 200 ms is superior by a factor of approximately 1.5 over sensitivity at an interstimulus duration of 700 ms.

II. EXPERIMENT II: FREQUENCY EFFECTS

In another test of the channel model, the importance of the frequency content of the profile is explored. Past work has shown that an optimal decision strategy provided by the Durlach *et al.* (1986) channel model effectively compares the mean level of the nonsignal components with the level of the component containing the signal. If the signal is an increment to one component of a standard, an optimal model predicts equal weighting of each nonsignal component (Berg and Green, 1990; Green, 1992). By Durlach *et al.*'s (1986) channel model, then, if the processing is the same and independent in all channels, and if all σ_i are equal, thresholds would be independent of the frequency distribution of the constituent components as long as the number of components in the profile is held constant. The intent of the follow-

ing experiment is to test this feature of the model by exploring the relative contribution of the components of the profile on the detection of a tonal signal. This was achieved by systematically varying the frequency distribution of the standard.

This experiment consists of two parts. First, sensitivity to a level increment (signal) of a single component of a 23-component equal-amplitude standard is evaluated at several different signal frequencies. This portion of the experiment replicates past work (e.g., the "profile bowl," Green and Mason, 1985; Bernstein and Green, 1987). In the second portion of experiment II, the frequency distribution of the standard is varied by introducing "gaps" in the profile. A gap is produced by removing six of the 23 tones that comprise the beginning standard. The constituent components are removed so as to provide a spectral profile with gaps placed either symmetrically on each side of the signal component or to one side or the other (higher or lower in frequency) with respect to the signal. The idea is that removal of more important components would reduce sensitivity to the added signal. The two portions of experiment II will be referred to as Experiments IIa and IIb.

A. Methods for experiment II

1. Stimuli

The standard was the sum of 23 equal-amplitude tones. The component tones were equidistant on a logarithmic frequency scale and ranged in frequency from 200 to 5000 Hz. On each presentation, the phase of each component was randomly chosen from a uniform distribution ranging from 0 to 2π rad. The signal to be detected was a pure tone added in-phase to one of the component tones.

For experiment IIa, which was the first completed, all 23 components were present. Five signal frequencies were tested: 349, 645, 1000, 1795, and 2785 Hz. These frequencies correspond to the 5th, 9th, 12th, 15th, and 19th component of the 23-component complex, respectively.

For experiment IIb, the standard was the sum of 17 equal-amplitude tones, ranging in frequency from 200 to 5000 Hz. The minimum intercomponent spacing remained the same as in the 23-component complexes (experiment IIa), however, various components were excluded, producing a spectral gap in the standard complex.

The three signal frequencies are 349, 1000, and 2785 Hz. These will be referred to as the 5th, 12th, and 19th components, respectively, in order to maintain consistency between these stimuli and the 23-component stimuli described in experiment IIa. Figure 3 displays signal-present power spectra for a 1000-Hz signal frequency. The 23-component complex is referred to as "full," as all tones in the stimulus are present. The 17-component complexes are labeled in the following manner: For the "above" condition, the six components immediately above the signal frequency are absent, and for the "below" condition, the six components immediately below the signal frequency are absent. It should be noted that the "below" condition does not exist when the signal is at the 5th component, and analogously, the "above" condition does not exist when the sig-

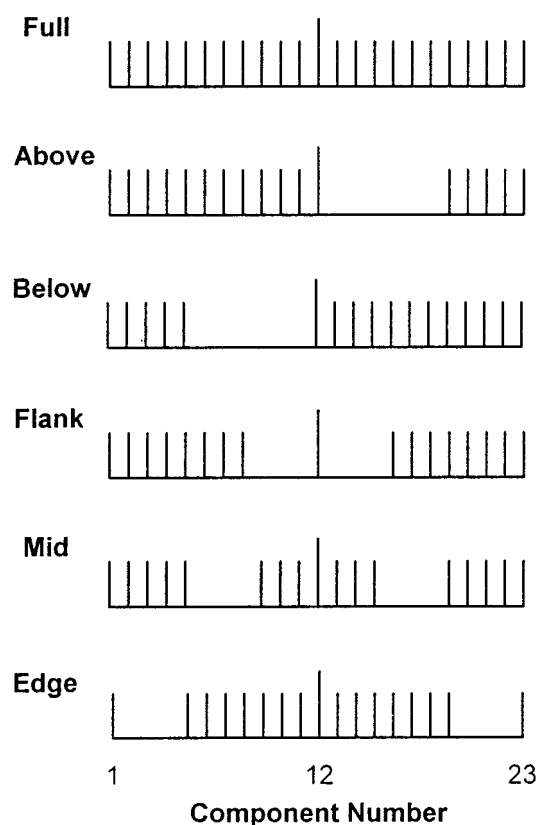


FIG. 3. Examples of signal-present power spectra are shown for a 1-kHz signal. The “full” condition consists of all 23 components. All other conditions are missing six components compared to the “full” condition. In the “above” condition, six components directly above the signal component are excluded, and in the “below” condition, six components directly below the signal component are removed. Three components directly above and three components directly below the signal component are removed to produce the “flank” condition. The “mid” and “edge” conditions only exist when the signal frequency is 1000 Hz. The “mid” condition is produced by removing the 6–8th and 16–18th components. The “edge” condition is produced by removing the 2–4th and 20–22nd components.

nal is at the 19th component. For the “flank” condition, the three components immediately above and below the signal component are absent.

For the 1000-Hz signal (12th component), two additional conditions tested are described below. Like the “flank” condition described above, the gaps for these additional conditions are symmetric about the signal frequency. For the “mid” condition, the 6–8th and 16–18th components are absent (i.e., midway between the edges and the middle of the profile). For the “edge” condition, the 2–4th and 20–22nd components are absent (i.e., at the edge of the profile).

2. Procedure

The stimuli were generated digitally and played via two channels of a 16-bit digital-to-analog converter (DAC; TDT DA1). The sampling rate was 20 kHz and the DAC output was low-pass filtered at 6 kHz (Stewart VBF 10M; attenuation skirts of approximately 85 dB/oct). Stimuli were presented diotically through Sennheiser HD 450 II earphones, and observers were seated in a single-walled sound-attenuated booth. The total duration of each stimulus was

100 ms, including 10-ms cosine-squared rise/decay ramps. The interstimulus interval duration was approximately 750 ms. The overall levels of the stimuli were randomly varied from the mean level of 50 dB SPL per component based on draws from a uniform distribution of ± 15 dB using a 0.1-dB gradation.

Unlike in experiment I, there was no danger of confusion about which stimulus contained the signal. Therefore a two-alternative forced choice paradigm was employed, and a two-down, one-up adaptive staircase procedure (Levitt, 1971) was used to estimate the signal level needed to achieve 71% correct. The signal to be detected was a tone added in-phase to one of the components of the standard, the “signal” component, and was as likely to occur in the first interval as the second. The strength of the signal is described as the amplitude of the signal relative to the amplitude of the component to which it was added—*sig re*: comp in dB. The initial signal level was set approximately 10 dB above the final threshold. The signal level was initially altered by 4-dB steps, and the step size was reduced to 2 dB upon the third reversal of the staircase. Observers responded via keyboard and were given feedback after each trial. Threshold estimates were measured in 50-trial sets, and the last even number of reversals, excluding at least the first three, were averaged to generate a threshold estimate. At least 15 threshold estimates were measured per condition, and the final estimate was obtained by averaging the last ten threshold estimates.

3. Observers and order of data collection

Three observers, ranging in age from 19 to 24 were tested.⁴ All observers had hearing thresholds within 15 dB of normal between 250 and 8000 Hz. Only observer 3 participated in both experiments I and II.

All observers completed experiment IIa first. Observers practiced at 1000 Hz, and data collection began after thresholds appeared to asymptote, which required a minimum of 4 h of practice. The order in which the different signal frequencies were tested was randomly chosen for each observer. Time allowed for tests of practice for only a few conditions. When repeated, these conditions showed no effects of practice. Thresholds reported reflect the average of the ten most recent threshold estimates.

The “full” thresholds reported in experiment IIb were taken from experiment IIa; that is, the conditions were not repeated. The remaining conditions of experiment IIb were completed as follows. The three signal frequencies (349, 1000, and 2785 Hz; the 5th, 12th, and 19th components) were completed in different orders for different observers, and for each signal frequency the different gap conditions (i.e., flank, mid, edge, below, and above) were blocked and completed in random order.

B. Results of experiment IIa

Figure 4 displays the results of experiment IIa. Thresholds are plotted as a function of signal frequency. Each panel shows the data for the individual observers, and error bars represent standard errors of the mean. Observers 3 and 4 show a “bowl” shape as the signal frequency is varied, with

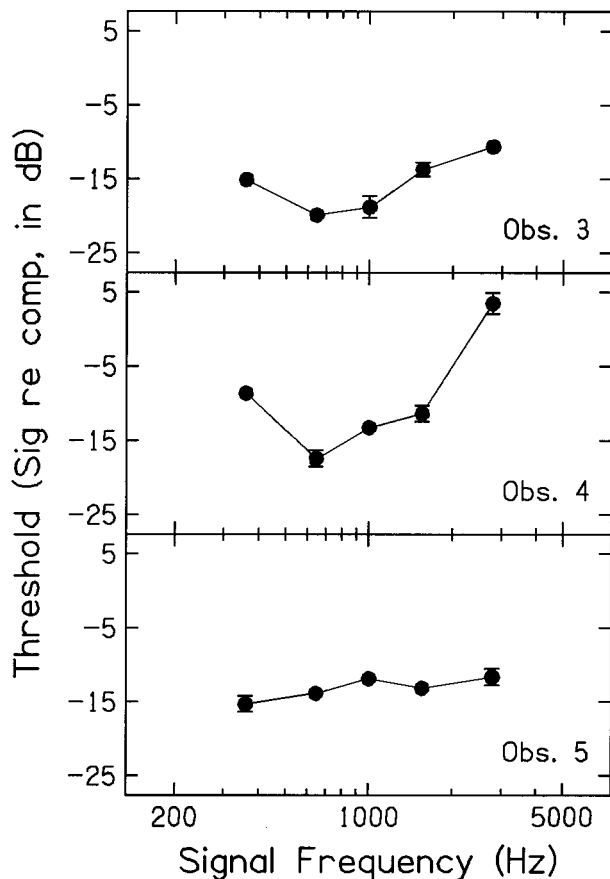


FIG. 4. Thresholds, as sig re: component, in dB, are plotted as a function of signal frequency. Each data point is the average of ten threshold estimates and error bars are standard errors of the mean.

the lowest thresholds at a signal frequency of 645 Hz. Observer 5 shows a more linear relationship between signal frequency and threshold. It should be noted that the threshold presented for observer 4 at 2875 Hz is not below what is expected if she had been using level cues, a value of 2 dB (Green, 1988b). The pattern of results shown in Fig. 4 is consistent with published work (Green and Mason, 1985; Green *et al.*, 1987).

C. Results of experiment IIb

Figure 5 shows average thresholds for the three observers. Error bars indicate the standard errors of the mean across observers. The different conditions tested are indicated at the bottom of the plot. Filled bars represent thresholds for an increment at the 12th component (1000 Hz), and striped bars represent thresholds for an increment at the 5th component (349 Hz). Due to large individual differences for the 19th component (2785 Hz), average thresholds are not shown. In addition, a threshold could not be obtained in the “below” condition for the 19th component for observer 4. The individual data are recorded in Table I.

On average the thresholds obtained using a 23-component standard are somewhat lower than those obtained when 17 components comprised the standard, although the magnitude of the difference is small (less than 2 dB). For the 17-component standard (Fig. 5, bars to the right of the

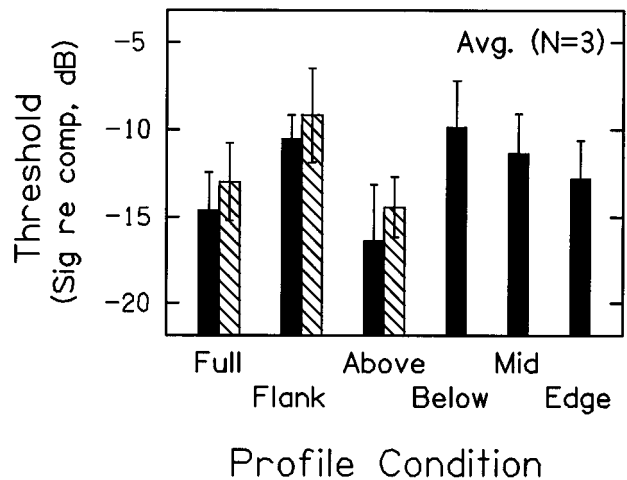


FIG. 5. Thresholds averaged across the observers for the different conditions are shown. Error bars are standard errors of the mean across three observers. The solid bars represent thresholds for the 1000-Hz signal (12th component), and the striped bars represent thresholds for the 349-Hz signal (5th component). Average thresholds for the 2785-Hz signal (19th component) are not presented due to sizable individual differences at this signal frequency.

“full” condition), thresholds appear to depend on the distribution of components as a function of frequency. The most notable effect is that thresholds in the “above” condition are always lower than the other thresholds obtained using the

TABLE I. Thresholds for each observer are indicated as signal re: component, in dB.

| | Observer 3 | Observer 4 | Observer 5 | Avg. |
|-----------------------------|----------------|----------------|----------------|----------------|
| Sig. at 359 Hz (No. 5) | | | | |
| full | -15.1 (0.6) | -8.6 (0.9) | -15.3 (1.1) | -13.0 (2.2) |
| flank | -11.7 (1.3) | -3.8 (1.3) | -12.1 (1.4) | -9.2 (2.7) |
| above | -16.1 (1.5) | -11.0 (1.3) | -16.3 (1.2) | -14.5 (1.7) |
| Sig. at 1000 Hz (No. 12) | | | | |
| full | -18.8 (1.5) | -13.2 (0.8) | -11.8 (0.9) | -14.6 (2.1) |
| flank | -12.1 (1.4) | -7.8 (1.3) | -11.6 (1.6) | -10.5 (1.3) |
| above | -22.5 (3.1) | -11.5 (1.1) | -15.2 (1.6) | -16.4 (3.2) |
| below | -15.2 (0.9) | -7.0 (0.9) | -7.4 (1.5) | -9.9 (2.7) |
| mid | -15.7 (1.4) | -9.8 (0.9) | -8.4 (1.1) | -11.3 (2.2) |
| edge | -16.9 (1.1) | -9.6 (1.1) | -11.9 (2.0) | -12.8 (2.1) |
| Sig. at 2785 Hz (No. 19) | | | | |
| full | -10.6 (0.5) | 3.5 (1.4) | -11.6 (1.1) | -6.2 (4.9) |
| flank | -6.4 (0.9) | 1.2 (1.1) | -9.0 (1.2) | -4.7 (3.1) |
| below | -11.5 (1.5) | *** | -6.6 (1.1) | |

other 17-component conditions (Table I). Less consistent across observers, the “flank” and “below” conditions tend to have higher thresholds than the other conditions (Fig. 5 and Table I).

The data of experiment IIb indicate that changes in the location of the gap within the 17-component standard have a modest influence on detection thresholds. Because removing components directly above the signal frequency is associated with the lowest thresholds, it would seem that the components directly above the signal act to “mask” the addition of the signal. These results suggest that observers are not weighing all elements of the profile equally. That conclusion is not predicted by the assumptions made here in conjunction with the channel model of Durlach *et al.* (1986); nor do the data of Green and Berg (1991) suggest differential weighing of information above and below the signal frequency.

In an effort to explore the possibility that “energetic” masking was the basis for the pattern of results obtained using the 17-component profile, excitation patterns were constructed for the different conditions tested. The excitation patterns were based on auditory filters as described by Glasberg and Moore (1990). For the stimulus levels used in the current experiment, the skirts of the auditory filters are approximately symmetric and, as a result, the excitation patterns suggest no explanation for the different effects of placing a gap directly above or directly below the signal frequency.

III. SUMMARY

The experiments described above were completed to provide additional tests of the Durlach *et al.* (1986) channel model. For the first experiment, the data show that sensitivity to changes in spectral shape is superior to changes in level, even when no rove is present to degrade level discrimination. A simple instantiation of the Durlach *et al.* (1986) channel model predicts that sensitivity would be identical for these two tasks at 0-rove. In an attempt to account for these differences, two models of central noise as described by Durlach *et al.* (1986) are explored. When the central noise acts after the formation of the likelihood ratio the effects of overall level variation are overemphasized, and the fits account for approximately 72% of the available variance. In contrast, when the central noise acts prior to the formation of the likelihood ratio, the fits to the data account for at least 95% of the available variance. For the current experiment, this model parallels previous models of central noise derived for the detection of increments in the level of tonal stimuli (Durlach and Braida, 1969).

In the second experiment, by measuring the detection of a level increment to one component of a multitone complex, the channel model was again evaluated. For the channel model, assuming that processing is the same in all channels and all σ_i are equal, systematically removing six components from a 23-component complex would yield the same threshold for all locations of a spectral “gap.” The results indicate, however, that sensitivity is the best in the conditions with spectral gaps located above the signal frequency. This may be due to the fact that when all components are present, the components directly above the signal in frequency are mask-

ing the addition of the signal. Because the location of the gap influences thresholds, it is possible that one or more of the assumptions incorporated into many tests of the channel model are inaccurate. Whatever the case, the current data are at odds with the relative weight estimates reported by Green and Berg (1991).

ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC 02012 from the National Institutes on Deafness and Other Communication Disorders, National Institutes of Health and University Research Foundation, University of Pennsylvania. The authors wish to thank Dr. Jacob Nachmias and Dr. Emily Buss for their comments on earlier drafts of this manuscript. Dr. Charlotte Reed, Dr. Louis Braida, and Dr. Robert Lutfi contributed to the further improvement of this manuscript as well.

¹Averages were across d' scores rather than percent correct. Occasionally obtained 100% correct scores were assigned a d' score of 3.8.

²Without loss of generality, it can be assumed that the central noise is not present in the shape discrimination task.

³The invertable transformation Durlach *et al.* (1986) employs generates random variables X_i such that X_1 is the sum of level estimates across all active channels, and X_k compares the level estimate in channel k against the average of level estimates in channels $(1, 2, \dots, k-1)$.

⁴One additional observer began the task, but quit midway through the experiment. Another observer was recruited, but did not begin the experiment as her thresholds failed to exceed the predictions of the energy model (a criterion adopted as part of the experimental protocol).

- Berg, B. G., and Green, D. M. (1990). “Spectral weights in profile listening,” *J. Acoust. Soc. Am.* **88**, 758–766.
- Berliner, J. E., and Durlach, N. I. (1972). “Intensity perception. IV. Resolution in roving-level discrimination,” *J. Acoust. Soc. Am.* **53**, 1270–1287.
- Bernstein, L. R., and Green, D. M. (1987). “Detection of simple and complex changes of spectral shape,” *J. Acoust. Soc. Am.* **82**, 1587–1592.
- Durlach, N. I., and Braida, L. D. (1969). “Intensity perception. I. Preliminary theory of intensity resolution,” *J. Acoust. Soc. Am.* **46**, 372–383.
- Durlach, N. I., Braida, L. D., and Ito, Y. (1986). “Towards a model for discrimination of broadband signals,” *J. Acoust. Soc. Am.* **80**, 63–72.
- Farrar, C. L., Reed, C. M., Ito, Y., Durlach, N. I., Delhorne, L. A., Zurek, P. M., and Braida, L. D. (1987). “Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises,” *J. Acoust. Soc. Am.* **81**, 1085–1092.
- Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notched-noise data,” *Hearing Res.* **47**, 103–138.
- Green, D. M. (1988a). “Auditory Profile Analysis: Some Experiments on Spectral Shape Discrimination,” in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowen (Wiley, New York), pp. 609–622.
- Green, D. M. (1988b). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).
- Green, D. M. (1992). “The number of components in profile analysis tasks,” *J. Acoust. Soc. Am.* **91**, 1616–1623.
- Green, D. M., and Berg, B. G. (1991). “Spectral Weights and the Profile Bowl,” *Q. J. Exp. Psychol.* **43A**, 449–458.
- Green, D. M., Kidd, G., and Picardi, M. C. (1983). “Successive versus simultaneous comparison in auditory intensity discrimination,” *J. Acoust. Soc. Am.* **73**, 639–643.
- Green, D. M., and Mason, C. R. (1985). “Auditory profile analysis: Frequency, phase, and Weber’s Law,” *J. Acoust. Soc. Am.* **77**, 1155–1161.

- Green, D. M., Onsan, Z. A., and Forrest, T. G. (1987). "Frequency effects in profile analysis and detecting complex spectral changes," J. Acoust. Soc. Am. **81**, 692–699.
- Kidd, Jr., G., Mason, C. R., and Hanna, T. E. (1988). "Evidence for sensory-trace comparisons in spectral shape discrimination," J. Acoust. Soc. Am. **84**, 144–149.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
- Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User's Guide* (Cambridge U.P., New York).
- Mason, C. R., Kidd, G., Hanna, T. H., and Green, D. M. (1984). "Profile analysis and level variation," Hearing Res. **13**, 296–275.

Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system

Robert P. Carlyon and A. Jaysurya Datta

MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

(Received 16 May 1996; revised 2 December 1996; accepted 4 March 1997)

A series of experiments compared the excitation produced in an auditory filter centered on 1100 Hz by two complexes, both of which consisted of harmonics 2–20 of a 100-Hz fundamental. When the components had a level of 69 dB SPL each, summing them in positive Schroeder phase produced substantially less forward masking of an 1100-Hz signal than when the components were summed in negative Schroeder phase. This difference decreased with decreases in overall masker level. Listeners also reported that the components of the positive-phase masker close to 1100 Hz were quieter than the corresponding components in the negative-phase masker. The data are explained using Kohlrausch and Sander's [J. Acoust. Soc. Am. **97**, 1817–1829 (1995)] finding that the response of an 1100-Hz auditory filter to the positive-phase complex shows marked peaks and dips, whereas that to the negative-phase complex does not. It is argued that the peaks in the response to the positive-phase masker are attenuated by fast-acting compression in the auditory system, thereby reducing the excitation produced by that sound. It is also argued that, compared to the power functions commonly used to model "excess masking" and the growth of loudness, the present data reflect greater compression at high levels but less compression at low levels. © 1997 Acoustical Society of America. [S0001-4966(97)06206-1]

PACS numbers: 43.66.Dc, 43.66.Mk [WJ]

INTRODUCTION

Auditory excitation can be loosely defined as the amount of neural activity, within a given frequency channel, that is elicited by a sound. Excitation in the human auditory system has been estimated using a number of psychophysical measures, perhaps the most common of which have been masking and loudness. Starting with the early data on the growth of loudness, the results of these experiments have suggested that the relationship between the external level of a stimulus and its internal representation in the auditory system is a compressive one (Zwicker, 1958; Hellman and Zwislocki, 1963; Zwicker and Scharf, 1965). More recently, the concept of a compressive nonlinearity in the auditory system has been used to model data on the combined effects of forward and backward maskers (Penner, 1980; Oxenham and Moore, 1994; Oxenham and Moore, 1995), on the detection of amplitude decrements (Moore *et al.*, 1996; Peters *et al.*, 1996), on temporal integration at and above threshold (Penner and Shiffrin, 1980; Florentine *et al.*, 1996), and on the masking by narrow bands of noise of simultaneously presented pure tones (van der Heijden and Kohlrausch, 1995).

Although no firm evidence has been provided concerning the possible site(s) of compression in human hearing, many of the more recent models have been inspired by the compressive nonlinearity observed in the peripheral auditory systems of other mammals. In this article we present new psychophysical evidence for compression in the auditory system, and show that, like the nonlinearity observed in the input–output functions of the basilar membrane ("BM": Sellick *et al.*, 1982; Ruggero *et al.*, 1992) and inner hair cells ("IHCs": Russell and Sellick, 1978; Meddis, 1988), this compression must be fast-acting, operating on a cycle-by-cycle basis. In addition, we argue that the amount of com-

pression applied to the stimulus varies with level in a manner qualitatively consistent with that observed in the responses of the BM and of IHCs. Most of our discussion will focus on these cochlear sites; however, the potential role of neurally based fast-acting compression will be considered in Sec. VI C.

One way in which fast-acting compression could affect the amount of auditory excitation produced by a sound is illustrated by Fig. 1, which shows two simulations of the output of a hypothetical auditory filter. The precise details of the simulations are described in Sec. VI B; for the purposes of the present discussion, it is sufficient to note that, although the outputs have the same power, they differ markedly in "peakiness."¹ If one calculates excitation by rectifying the filter output, compressing it, and averaging the resulting values over time (Meddis and Hewitt, 1991; Patterson, 1994; Patterson *et al.*, 1995; Carlyon, 1996), then the peaks in the filter output shown in Fig. 1(a) will be attenuated, and the resulting excitation will be less than for the output shown in Fig. 1(b). Consequently, models which include such fast-acting compression predict that excitation, calculated from the compressed output of a given auditory filter, can differ between two sounds with identical power spectra; this difference will occur whenever the filter outputs to the two sounds differ markedly in peakiness (Meddis and Hewitt, 1991; Patterson *et al.*, 1995). Such a prediction is, of course, never made by models which initially calculate the power spectrum of the input (e.g., Zwicker, 1958; Zwicker and Scharf, 1965; Moore and Glasberg, 1983, 1996), even if a compression is applied at a later stage.

A recent article presented evidence that sounds with identical power spectra can indeed differ in the excitation which they produce. Carlyon (1996) showed that a tone hav-

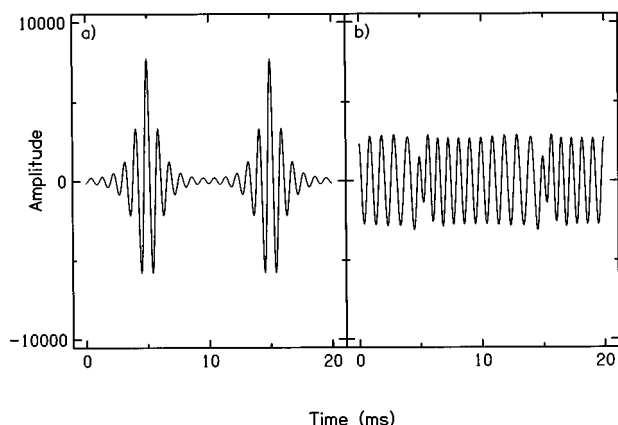


FIG. 1. Two cycles of the “minimally” and “maximally” peaked waveforms used in the simulated filter outputs, as described in Sec. VI B 2.

ing a “ramped” envelope (Patterson, 1994) more effectively masked a signal above its carrier frequency (f_c) than did a tone having a “damped” envelope, but that the opposite was true for signals below f_c . This occurred despite the fact that the two maskers were simply time-reversed versions of each other, and therefore had the same long-term power spectrum. Carlyon showed that a modification of a cochlear model (Giguère and Woodland, 1994) made predictions which were in qualitative agreement with the pattern of results observed: The model produced filter outputs which were more peaked in response to the damped than to the ramped tone for filters having center frequencies (CFs) above f_c , and more peaked in response to the ramped tone for filters with CFs below f_c . Because the model also included a fast-acting compression, the peaked filter responses produced less excitation than those with flat envelopes, resulting in predicted excitation patterns which reflected the masking asymmetry seen in the data. However, Carlyon did not present any experimental evidence that the shape of the filter outputs actually differed in this way. The present study provided a more direct test of the idea that fast-acting compression reduces excitation when the response of an auditory filter contains marked peaks and dips, compared to the case where it has a flat envelope.

The experiments reported here used both forward-masking and loudness-judgement techniques to estimate the excitation at the output of an auditory filter, centered on 1100 Hz, in response to two sounds. The sounds, which had identical power spectra, consisted of 19 equal-amplitude harmonics of a 100-Hz fundamental (F_0), summed in either “negative” or “positive” Schroeder phase (Schroeder, 1970). The experiments exploited Kohlrausch and Sander’s (1995) finding that the threshold for a brief 1100-Hz signal varied by as much as 17 dB when its temporal position was stepped through one period of the positive-phase masker. They concluded that the amplitude at the output of an 1100-Hz auditory filter varied markedly throughout each 10-ms period of that masker, but that this was not the case for the negative-phase masker, whose corresponding “masking period pattern” (MPP) was flat. Because the filter outputs had equal power, this leads to the prediction that, if excitation is based on a compressed version of filter outputs, then the positive

phase complex should produce less excitation than the negative phase complex.

I. EXPERIMENT 1

A. Method and rationale

All of the forward-masking experiments described here were based on the assumption that, if the positive-phase complex produces less excitation than the negative-phase complex in an auditory filter centered on the signal, then, at a given masker-signal delay, it should produce less forward masking. Although there is some debate concerning the auditory processes responsible for forward masking, this general assumption seems reasonable and is widely made in psychoacoustic studies both of compression (Penner, 1980; Oxenham and Moore, 1994, 1995) and of other auditory processes such as frequency selectivity (e.g. Moore, 1978). In Sec. VI C we discuss the implications of one specific explanation of forward masking, adaptation, for the nature and amount of compression needed to account for the results of our forward-masking experiments.

The forward maskers used in experiment 1 were the positive and negative Schroeder phase complexes described by Kohlrausch and Sander (1995). Each consisted of harmonics 2–20 of a 100-Hz fundamental (F_0), presented at a level of 69 dB SPL per component. The starting phase of the n th harmonic equaled $\pi n(n-1)/19$ for the positive-phase complex and $-\pi n(n-1)/19$ for the negative-phase complex. The signals were 1100-Hz sinusoids gated on and off with 5-ms raised-cosine ramps (no steady state) and presented 5 ms (zero-voltage points) after the masker had been turned off. Two factors need to be considered in order to ensure that any difference in forward masking produced by the two complexes are not due to short-term fluctuations in the maskers or in the auditory filter responses to them. The first arises from the variation in instantaneous frequency which occurs throughout each period of the maskers. Kohlrausch and Sander (1995, Fig. 2) have shown that, 5 ms after the start of each period, the two maskers have the same instantaneous frequency of 1100 Hz. This would be an ideal time to turn the masker off; however, our maskers were gated with 5-ms raised-cosine ramps, and it is unclear whether the instantaneous frequencies should be equated at the beginning, middle, or end of the offset ramp. We chose to do all three, and therefore used masker durations (between the zero-voltage points) of 395, 397.5, and 400 ms; for these three durations the offset ramp spanned times from, respectively, 0–5, 2.5–7.5, and 5–10 ms after the start of the final masker period. The phase of the components at the beginning of each masker was the same for all three masker durations. A second reason for using more than one masker duration arises from the highly peaked MPP seen for the positive-phase masker (Kohlrausch and Sander, 1995; Carlyon and Datta, 1997). We wanted to control for the possibility that, by turning the masker off at the end of a valley in the output of the 1100-Hz auditory filter, we might have extended the “internal” masker-signal delay in the positive-phase condition. This could have decreased the amount of masking observed in that condition, and, if this were the

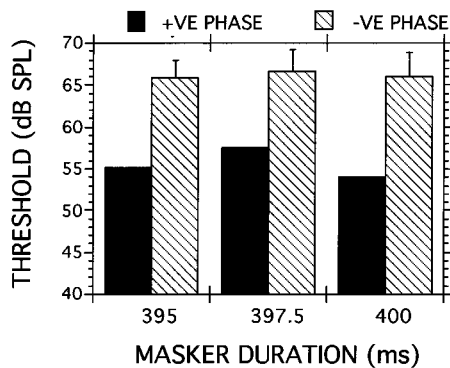


FIG. 2. Thresholds for a brief 1100-Hz signal presented 10 ms after the offset of the positive-phase (solid bars) and negative-phase (hatched bars) complexes, for three masker durations. Error bars were calculated by taking the difference between each listener's "positive phase" and "negative phase" thresholds and then calculating the standard error across listeners.

case, thresholds with the positive-phase masker should vary markedly over the three masker durations used here.

The maskers and signals were generated digitally and played out through separate channels of a CED 1401 laboratory interface at a sampling rate of 20 000 Hz and with 12-bit resolution. They were low-pass filtered (Kemo VBF25.01, attenuation rate=135 dB/oct; cutoff=8600 Hz), attenuated, and fed to separate inputs of one channel of a headphone amplifier. Four listeners, each with absolute thresholds within 15 dB of the 1969 ANSI standard, sat in a double-walled sound-attenuating booth and listened through one earpiece of a Sennheiser HD414 headset. Thresholds were estimated using a two-down one-up adaptive procedure (Levitt, 1971), with feedback being provided at the end of each 2IFC trial. The signal level was decreased by 2 dB after every two correct responses and increased by the same amount after every incorrect response. Each change from increasing to decreasing level and vice versa defined a turnpoint. Each run terminated after 16 turnpoints, and the estimate for that run was obtained by averaging the signal levels at the last 12 turnpoints. All data presented here were obtained from the mean of six runs per listener.

B. Results

The pattern of results was similar for the four listeners, and so mean data are shown in Fig. 2. The signal threshold was about 10 dB lower when presented after the positive-phase complex than after the negative-phase complex. This difference occurred for all three masker durations, indicating that it was dependent neither on the instantaneous frequency at the end of the masker nor on whether the masker was turned off at a peak or at a valley of the auditory filter response. The results demonstrate not only that the filter outputs are compressed, but also that the compression is fast-acting. Clearly, if it operated on the power of the filter output calculated using a window of more than 10 ms (the period of the masker), then the same gain would be applied to the filter responses to both complexes, and no difference in masking would occur.

II. EXPERIMENT 2: EFFECT OF SENSATION LEVEL

A. Rationale and method

Animal experiments indicate that the compression observed in the responses of the BM and of IHCs to sound is reduced at low levels (Russell and Sellick, 1978; Sellick *et al.*, 1982; Ruggero *et al.*, 1992). Thus, if the difference in forward-masking effectiveness observed in experiment 1 is mediated by the compression observed in the cochlea, we would expect it to become less pronounced at low masker levels. A second reason for measuring the level dependence of the effect is simply to observe the range of conditions under which the calculation of the power spectrum leads to an inaccurate measure of excitation.

The experiment started by measuring the quiet threshold of the 10-ms 1100-Hz signal for each of the five listeners who took part, using the same procedure as in experiment 1. The method used for the main part of the experiment was similar, except that the signal level was held constant and the masker level varied adaptively. One drawback of the signal-varying method used in experiment 1 arises from the decay of forward masking, which is faster for more intense stimuli; a consequence of this is that two maskers differing in level by a given amount may produce a similar difference in simultaneous masked thresholds, but a much smaller difference in forward masking (Jesteadt *et al.*, 1982). Hence, the difference in signal threshold produced by two maskers in forward masking may underestimate the difference in excitation that they elicit. By measuring the level of each complex necessary to mask a fixed signal, we were able to avoid the influence of the decay of forward masking on the results (Moore and Glasberg, 1981). This was done using a masker duration of 397.5 ms, a 5-ms masker-signal gap, and signal levels of 10, 20, and 30 dB SL. For two listeners, measurements were also made with a signal level of 40 dB SL.

B. Results and discussion

Listeners RC, CL, NM, PM, and GN had quiet thresholds for the 10-ms 1100-Hz signal of 25.4, 18.6, 17.5, 26.1, and 26.6 dB SPL, respectively. The open symbols in Fig. 3(a) show the levels (in dB per component) of the 397.5-ms complexes necessary to mask the signal, as a function of its sensation level. Although the positive-phase masker (squares) is less effective than the negative-phase masker (triangles) at high overall levels, this difference disappears at low levels. The dependence of the effect on masker level is further illustrated in Fig. 3(b) which shows the difference in level between the two maskers at signal threshold, as a function of the level of the negative phase masker. The figure shows that the two maskers are equally effective at levels up to about 30 dB SPL/component. For four out of five listeners the difference increases smoothly as level is increased further, and reaches a maximum of between 10 and 15 dB. The exception is listener GN, who shows no difference when the negative-phase masker has a level of 40 dB/component, but a 26-dB difference when it has a level of 54 dB/component.

The results of experiment 2 are consistent with the idea that the peripheral auditory system is less compressive at low than at moderate-to-high levels, as observed in the responses

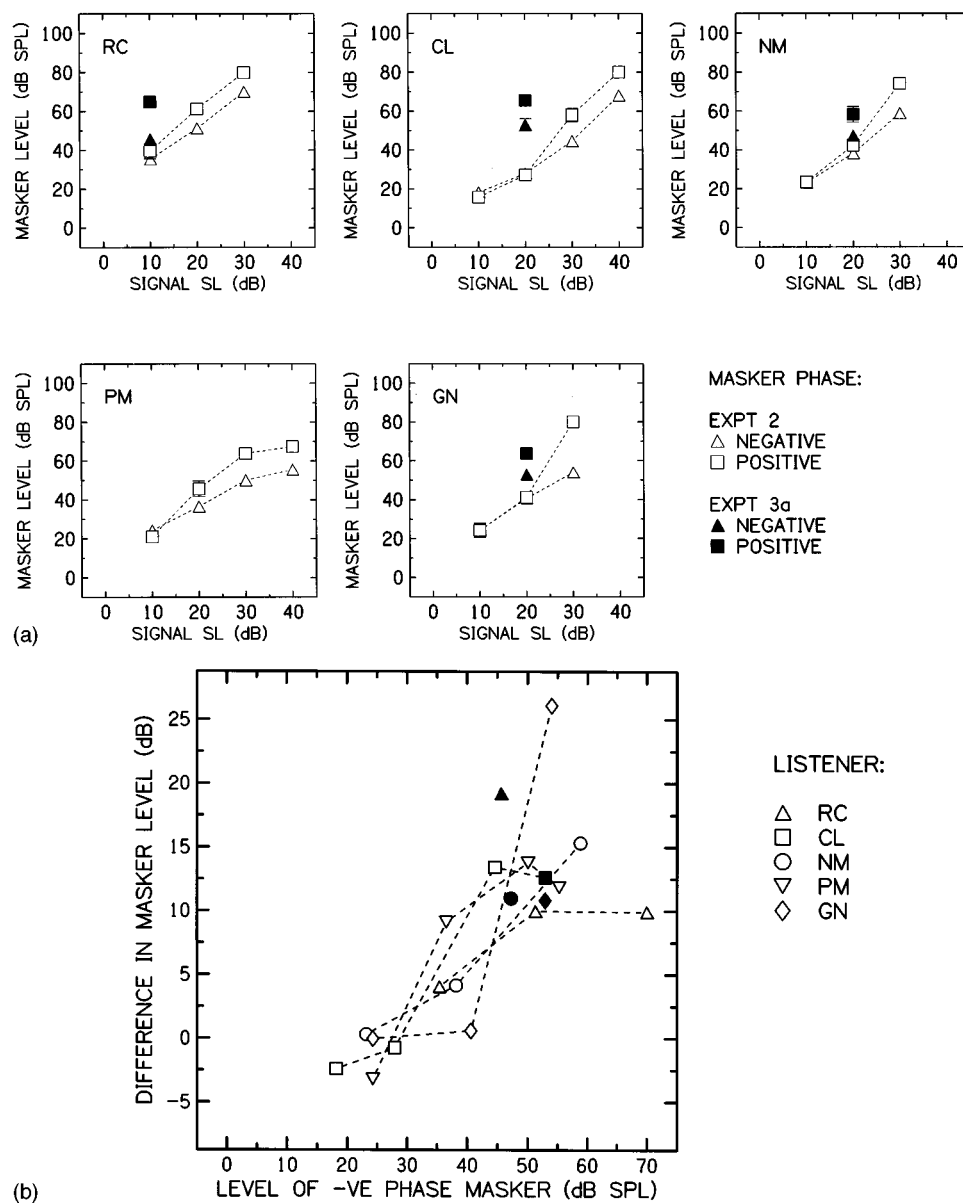


FIG. 3. (a) The levels per component of the positive-phase (squares) and negative-phase (triangles) complexes necessary to mask a subsequent 1100-Hz tone burst presented at 10, 20, 30, or 40 dB above its threshold in quiet. Open symbols were obtained with a masker duration of 397.5 ms. Filled symbols were obtained with a masker duration of 27.5 ms. (b) The data of (a) are replotted to show the difference between the levels of the positive- and negative-phase complexes necessary to mask a given signal, as a function of the level of the negative-phase masker.

of the mammalian BM and IHCs. The next two sections evaluate alternative explanations for the level dependence seen here.

III. EFFECT OF MASKER DURATION

A. Negative- and positive-phase maskers having short durations

1. Rationale and method

A previous suggestion that peripheral compression is reduced at low levels was made by Penner and Shiffrin (1980). They measured forward masking produced by pairs of noise bursts, which did not temporally overlap and which were separated from the signal by different masker-signal delays. The levels of the two bursts were adjusted so that, when presented individually, each masker was equally effective as

the other. When the two maskers were combined, threshold increased by more than 3 dB, and, importantly, the amount of this “additional,” or “excess” masking increased with increasing level. They concluded not only that the maskers were compressed by the peripheral auditory system but that, relative to a power function, the compression was reduced at low compared to high levels. However, Humes and Jesteadt (1989) have shown that Penner and Shiffrin’s data can be modeled using a simple power-law compression, if one assumes that, prior to compression, a fixed-level internal noise is added to the stimulus. An important prediction of this explanation is that the size of the effect should vary with the level of the signal, rather than that of the masker. In order to determine whether such an explanation can account for our data, experiment 3a repeated experiment 2 at a low signal level but with the masker duration reduced to 27.5 ms. Pre-

vious data (Zwislocki *et al.*, 1959; Kidd and Feth, 1982; Kidd *et al.*, 1984; Carlyon, 1988) suggest that this should raise the levels of both complexes necessary to mask the signal, thereby allowing us some control of the masker level whilst keeping the signal level constant. The experiment used a single signal level, at which the 397.5-ms positive- and negative-phase maskers in experiment 2 were equally effective. This level was selected separately for each listener, and was 10 dB SL for listener RC and 20 dB SL for listeners CL, NM, and GN. The prediction was that, if compression increases with increases in masker level, then this manipulation should introduce a difference between the positive- and negative-phase conditions.

2. Results

The filled symbols in Fig. 3(a) show that, as expected, reducing the masker duration increased the levels of both maskers necessary for them to mask the signal. More importantly, Fig. 3(a) and (b) shows that it introduced a substantial difference between the effectiveness of the two maskers, indicating that the level of the masker, rather than that of the signal, is important for the effect. This is inconsistent with an explanation based on a fixed-level internal noise being added to the stimulus.

B. Effect of masker duration at different levels and masker-signal delays

1. Rationale

Many models of forward masking (e.g., Penner and Shiffrin, 1980; Oxenham and Moore, 1994) account for forward masking by assuming that the masker and signal are subjected to a temporal weighting function which applies increasing amounts of attenuation to portions of the masker more and more remote in time from the signal. If the weighting function were the only transformation applied to the signal, then such models would predict that increases in masker duration should have only a small effect on signal thresholds. This effect would be smaller than that predicted from the increase in masker energy, because the “extra” portions of the masker would be attenuated by the temporal weighting function. However, by assuming that the auditory system also compresses the input, it is possible to account for somewhat larger effects of masker duration observed (Penner, 1979; Kidd *et al.*, 1984; Carlyon, 1988; Oxenham and Moore, 1994; see also our Fig. 3). This suggests that, if compression is reduced at low levels, then the effect of decreasing masker duration should also be reduced. Again, such a finding might be explained by assuming that a fixed internal noise is added to the stimulus, but only if the size of the effect varies with the level of the signal rather than that of the masker. Therefore, experiment 3b examined the effect of masker duration, while independently manipulating the levels of the maskers and signals. Unlike the experiments described so far, only negative-phase complexes were used.

2. Method

One condition involved measuring the level of the negative-phase masker, with durations of 27.5 and 397.5 ms, necessary to forward mask a 10-ms 1100-Hz sinusoid having

a sensation level of 20 dB. As before, the masker-signal delay was 5 ms. In fact, data at the two masker durations in this condition had already been obtained for the three listeners (CL, NM, and GN) in experiments 2 and 3a, and these data were re-used here. In the second condition the signal level was reduced to 10 dB SL. The prediction was that this would reduce the level of the complex necessary to mask the signal, and, if compression decreases at low levels, would also result in a reduced effect of masker duration. To control for the reduced signal level, this condition was re-run at a signal level of 10 dB SL but with the masker-signal delay increased to 20 ms; listener GN was also tested at an additional delay of 50 ms. This was expected to increase the masker level, and therefore also the effect of masker duration, while holding the signal level constant.

3. Results

The results of experiment 3b are shown in Table I, parts (a) and (b) of which show the data obtained with a masker-signal gap of 5 ms. Although there is considerable variability across listeners, the effect of masker duration (bold type) was reduced substantially by a 10-dB decrease in signal level for all listeners. This effect was restored by increasing the masker-signal gap to 20 ms [part (c)] for listeners CL and NM and to 50 ms [part (d)] for listener GN. Thus the effect of masker duration increased with masker level, even when the signal level was held constant. This is consistent with a reduction in compression at low levels, but not with the effects of masker duration being reduced at low levels by an additive, fixed-level internal noise.

IV. EFFECT OF MASKER LEVEL ON MASKING PERIOD PATTERNS

A. Rationale and method

The results of experiments 2 and 3a clearly show that the difference in masking effectiveness of positive and negative phase complexes disappears at low masker levels, and that this level dependence is not due to the low sensation level of the signal. Although the data are consistent with a reduction in compression at low levels, it is also possible that the outputs of auditory filters tuned to 1100 Hz did not differ in “peakiness” at these low levels. This is feasible, given that a very peaky output would require the interaction of a large number of components in the filter passband, and that auditory filters are narrower at moderate than at high levels (Glasberg and Moore, 1990; Rosen and Baker, 1994). To test this explanation, we measured MPPs for the two complexes at two masker levels. The higher level was always 69 dB SPL/component; the lower level was chosen separately for each listener on the basis of the results of experiment 2, such that the two complexes were equally effective forward maskers. The selected levels (with the level at threshold of the positive and negative phase maskers of experiment 2 in parentheses) were: 37.4 (39.4, 35.4) dB SPL/component for RC; 27.5 (27.1, 27.9) dB for CL; 23.3 (23.5, 23.1) dB for NM; and 40.9 (41.2, 40.6) dB for GN.

TABLE I. Level of 397.5- and 27.5-ms negative-phase maskers necessary to forward-mask an 1100-Hz signal at the signal levels and masker-signal delays (Δt 's) shown. Standard errors are shown in parentheses. The difference between the masker levels for the two masker durations are shown for each condition in bold italics.

| | | CL | NM | GN |
|------------------------|--------------------------|-------------|-------------|-------------|
| 20 dB SL | | | | |
| (a) $\Delta t = 5$ ms | $T = 397.5$ ms | 27.9 (1.1) | 38.2 (1.0) | 40.6 (0.8) |
| | $T = 27.5$ ms | 52.9 (3.1) | 47.3 (2.3) | 52.9 (1.9) |
| | <i>difference</i> | 25.0 | 9.1 | 12.3 |
| 10 dB SL | | | | |
| (b) $\Delta t = 5$ ms | $T = 397.5$ ms | 21.7 (1.2) | 23.2 (1.3) | 26.8 (0.6) |
| | $T = 27.5$ ms | 27.7 (2.8) | 22.3 (2.4) | 26.1 (2.3) |
| | <i>difference</i> | 6.00 | -0.9 | -0.7 |
| (c) $\Delta t = 20$ ms | $T = 397.5$ ms | 26.1 (2.3) | 29.5 (0.7) | 31.9 (1.0) |
| | $T = 27.5$ ms | 56.2 (4.1) | 39.0 (2.4) | 34.4 (1.7) |
| | <i>difference</i> | 30.1 | 9.5 | 2.5 |
| (d) $\Delta t = 50$ ms | $T = 397.5$ ms | ... | ... | 37.9 (1.1) |
| | $T = 27.5$ ms | ... | ... | 49.6 (2.8) |
| | <i>difference</i> | ... | ... | 11.7 |

As in the study by Kohlrausch and Sander (1995), the signal was a 5-ms 1100-Hz sinusoid with two 2.5-ms raised-cosine ramps and no steady state. For each masker and at each level, thresholds were measured for signals starting 152, 154, 156, 158, and 160 ms (zero-voltage points) after the onset of the masker. The masker duration was always 400 ms and the signal was always added in phase to the 1100-Hz component of the masker. The adaptive procedure was the same as in experiment 1.

B. Results and discussion

The open symbols in Fig. 4 show the MPPs obtained with a masker level of 69 dB SPL/component. The MPP for the negative-phase masker (triangles) is generally flat, with only small maxima and minima occurring at delays which differ across listeners. In contrast, the MPPs for the positive-phase masker (squares) show an average difference of about 18 dB between the highest and lowest points, which always

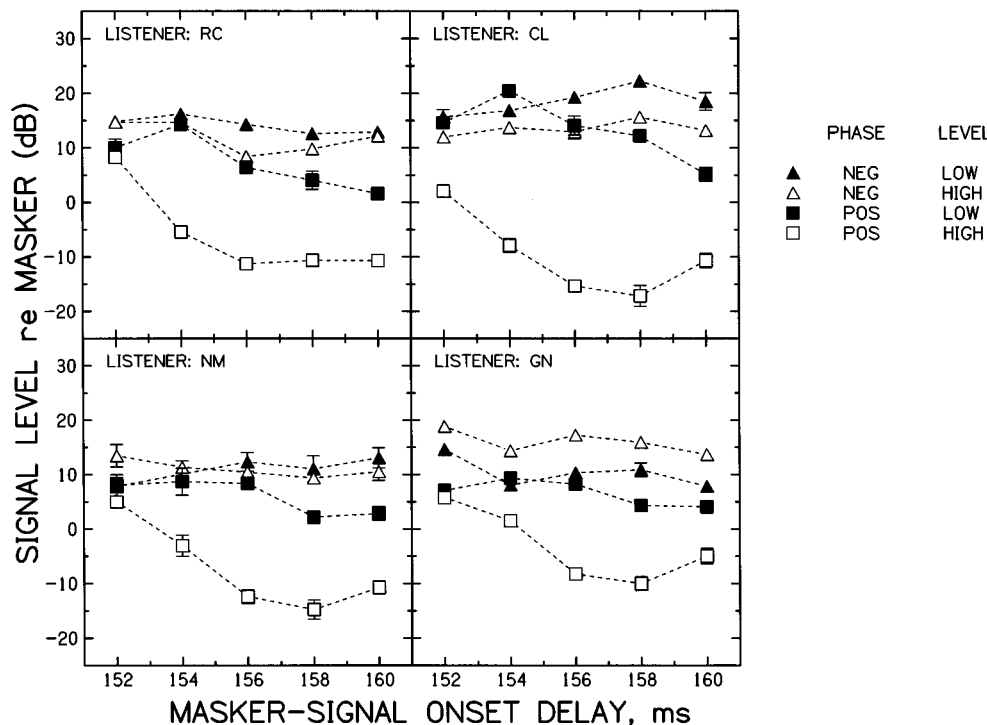


FIG. 4. Open symbols show masking period patterns obtained for the positive-phase (squares) and negative-phase (triangles) maskers, presented at a level of 69 dB/component. The solid symbols are for a lower masker level which was selected separately for each listener. This level, expressed in dB SPL/component, was 37.4, 27.5, 23.3, and 40.9 for listeners RC, CL, NM, and GN, respectively. Error bars show plus and minus one standard error.

occurs at masker-signal delays of 152 and 158 ms, respectively. These results agree well with similar measurements made by Kohlrausch and Sander (1995, Fig. 5) at a level of 63 dB component.

The filled symbols show the MPPs obtained at the lower level. Those for the negative-phase masker (triangles) are similar to those obtained at the higher level: Thresholds vary by between 3.6 and 6.8 dB over each listener's MPP, and the locations of the maxima and minima vary across listeners. For two listeners, GN and NM, the positive-phase MPPs (squares) at this level are also flat, unlike those obtained at the higher level, and show threshold variations of only 5.3 and 6.0 dB, not substantially greater than those seen with the negative phase masker. For these listeners, then, the absence of a forward-masking difference at low levels in experiment 2 can be largely accounted for by the shapes of the corresponding MPPs. However, this is probably not true for listeners RC and CL, whose positive-phase MPPs show threshold variations of 12.7 and 15.3 dB, respectively. For these listeners, the positive-phase MPPs (squares) are substantially more modulated than for the negative-phase complex (triangles), even at a level where the two complexes were equally effective forward maskers (experiment 2). Overall, it seems reasonable to conclude that, although an 1100-Hz auditory filter produces a flatter response to the positive-phase complex at low rather than at high levels, this can only partially account for the level dependence of the forward-masking effect seen in experiment 2. In this regard, it is also worth noting that the effects of masker duration obtained in experiment 3b, which we also attribute to compression, cannot be explained by a decrease in the peakiness of filter outputs at low levels. This is because experiment 3b used only negative-phase maskers, whose MPPs do not change shape with level. A discussion of the effect of masker level on the shape of the MPP, together with more detailed data on the topic, can be found in a companion article (Carlyon and Datta, 1997).

V. LOUDNESS JUDGMENTS

A. Rationale and method

Experiments 1–3 used forward masking to estimate the excitation produced by the two Schroeder-phase complexes. Another way of comparing the auditory excitation produced by two sounds is to ask listeners which is louder. The prediction is that, if fast-acting compression attenuates the large peaks in the responses of filters in a given frequency region to the positive-phase complex, then the “specific loudness” (cf. Zwicker, 1958; Zwicker and Scharf, 1965) in that region should be less than when the filters are stimulated by the negative-phase complex. Unfortunately, we cannot measure the specific loudness by presenting only those components of each complex with frequencies close to 1100 Hz: The auditory filter is rather broad at the high levels used here (Glasberg and Moore, 1990; Rosen and Baker, 1994), and the peaked MPP observed for the positive-phase complex requires the interaction of a large number of components within its passband (Carlyon and Datta, 1997). Thus, if we presented only a few components, the filter output would not

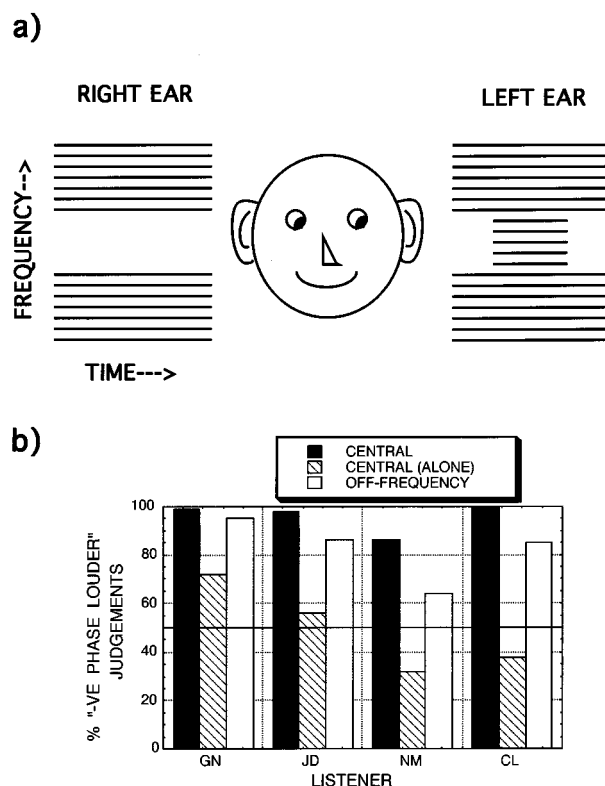


FIG. 5. (a) A schematic of the stimulus presentation for half of one trial in the main condition of experiment 4. The two halves of each trial were identical except for the phase in which the components were summed. The solid bars in part (b) show the percentage of trials on which listeners judged the central components of the negative-phase complex to be louder than the corresponding components of the positive-phase complex. The hashed bars show the results of the loudness comparison when the diotic “off-frequency” components were removed. The shaded bars show data for the condition where listeners compared the loudness of the diotic off-frequency components, in the absence of the central components.

show marked peaks and dips and would be compressed less than if all components were present. Conversely, the use of a large number of components would excite auditory filters having center frequencies remote from 1100 Hz, where the MPPs do not always differ in peakiness between the two complexes (Kohlrausch and Sander, 1995). Our solution was to use the technique shown schematically in Fig. 5(a) and described below. It was developed so that listeners could estimate the loudness in a fairly restricted frequency region, while allowing a large number of harmonics to interact in filters tuned to that region.

On half of the trials in the main condition, the harmonics comprising the positive-phase complex were presented in the stimulus configuration shown in Fig. 5(a), followed, 500 ms later, by the harmonics of the negative-phase complex presented in the same way. On the other half, the order of the two sounds was reversed. All harmonics except those with frequencies between 900 and 1300 Hz inclusive were presented diotically for 750 ms; the remaining “central” harmonics were turned on 187.5 ms later, had a duration of 375 ms, and were presented to the left ear only. Listeners indicated which of the two shorter sounds, heard in the left ear, seemed louder. No feedback was given. In a second condition listeners compared the loudness of the five central com-

ponents of the two sounds in the absence of the diotic “off-frequency components.” The situation was reversed in a third condition, in which listeners compared the loudness of the two diotically presented off-frequency complexes, with the central components absent.

In all conditions stimuli were digitally generated at a sampling frequency of 10 000 Hz and low-pass filtered (Kemo VBF 25.01) at 4300 Hz. All harmonics were turned on and off with 5-ms raised-cosine ramps and had a level of 69 dB each. Three listeners made 100 judgments in each of the three conditions. Listener NM also made 100 judgments in the first two conditions, but only made 80 judgments in the third, “off-frequency” condition.

B. Results and discussion

Figure 5(b) shows, for each condition, the percentage of trials in which listeners judged the negative-phase complex to be louder than the positive-phase complex. The solid bars show that in the main condition, where listeners judged the central components in the presence of the off-frequency components, the negative-phase complex was judged louder between 86% and 100% of the time.² This is highly significant, given that the upper 99% confidence limit for chance performance corresponds to a score of 63%. It is consistent with the idea that the specific loudness arising from auditory filters close to 1100 Hz is reduced by compression when stimulated by the positive-phase complex. The importance of having the off-frequency components present is illustrated by the hashed bars, which show that, when they were absent, the trend was either reduced (listeners GN and JD) or reversed (NM and CL). This is consistent with the finding that, when only the five central components are present, MPPs for the positive- and negative-phase maskers do not differ markedly in peakiness (Carlyon and Datta, 1997).

The shaded bars in Fig. 5(b) show that the negative-phase “off-frequency” components were judged to be louder than their positive-phase counterparts, although this trend was weaker than for the central components presented “in context” (solid bars). This is consistent with Kohlrausch and Sander’s (1995) finding that, although the MPPs for the positive- and negative-phase maskers differ most in peakiness for signals close to 1100 Hz, significant differences are observed for signal frequencies as low as 600 Hz and as high as 1600 Hz.

In summary, when considered with the results of other studies (Kohlrausch and Sander, 1995; Carlyon and Datta, 1997), the results of experiment 4 show that the loudness resulting from the outputs of a given range of auditory filters is affected by the peakiness in the response of those filters, with more peaked responses leading to lower loudness.

VI. DISCUSSION

A. Comparison with previous studies

1. Evidence for level-dependent compression

Perhaps the first indication that compression might be reduced at low levels, relative to that predicted by a power law, comes from data on the growth of loudness. Using a magnitude-estimation technique, a number of authors (e.g.,

Hellman and Zwislöcki, 1963) have shown that loudness appears to increase as a compressive power function of intensity at moderate-to-high levels, but increases more steeply at low levels. However, these and other data have been effectively modeled using a power-law compression covering the entire dynamic range, combined with an additive internal noise with a fixed level (Humes and Jesteadt, 1991). Of course, this does not prove that loudness-growth data do not reflect a reduction in compression at low levels, but it does show that the data can be successfully modeled without assuming that such a reduction exists.

A more recent suggestion that loudness judgements may reflect a compression which deviates from a power law comes from a study by Florentine *et al.* (1996), who measured the level of a long-duration tone or noise which was judged to be equally loud as a shorter stimulus of the same type. Temporal integration, defined as the difference between the levels of the long and short sounds, was greater for both types of stimulus at intermediate rather than at either high or low levels. They suggested that the auditory system is most compressive at intermediate levels, so that short stimuli would have to undergo a large increase in “external” level in order to produce the change in internal level necessary for their loudness to equal that of longer stimuli. Although it is possible that the reduction in temporal integration at low levels may be due to a fixed-level internal noise, similar to that proposed by Humes and Jesteadt (1991), no such explanation can account for the reduction observed at very high levels. However, Florentine *et al.* also obtained an excellent fit to their data using Zwislöcki’s (1965) modified power function, which does not include a level-dependent compression, by assuming that loudness grows more steeply for brief than for long stimuli.

Another source of evidence for auditory compression comes from experiments in which two equally effective non-simultaneous maskers are combined. A number of studies have shown that, provided the two maskers do not overlap in time, the combination produces more masking than predicted by the simple linear summation of their “masking effectiveness.” Once factors such as “off-time listening” (Robinson and Pollack, 1973) have been controlled for, such studies provide convincing evidence for the existence of compression in the auditory system. As discussed in Sec. IV A 1, Penner and Shiffrin (1980) showed a decrease in excess masking at low signal levels, which they attributed to a reduction in compression, but which Humes and Jesteadt (1989) succeeded in modeling using a simple power law combined with an additive internal noise. Similarly, Oxenham and Moore (1995) presented a model which predicts a reduction in “excess” masking at low signal levels, again by virtue of an additive internal noise. In contrast, we have been able to rule out such an explanation by demonstrating that the sizes of the effects observed here vary with the level of the masker, rather than with that of the signal. This is not possible when modeling excess masking, which, as Oxenham and Moore (1995) have pointed out, is determined by the compression applied to the signal and not by that applied to the masker.³ This makes it hard to tell whether the reduction in excess masking at low signal levels is due to a genu-

ine decrease in compression, to a fixed internal noise, or some combination of the two.

2. Evidence for fast-acting compression

In 1980, Penner and Shiffrin attempted to resolve the discrepancy between the long time constant of temporal integration for the detection of sounds and the much shorter time constants observed in measurements of temporal acuity. They showed that a power-law compression combined with an exponential integrator having a time constant of only 2–4 ms could account for temporal integration over hundreds of milliseconds. However, it should be noted that there have also been attempts to resolve this discrepancy without applying compression, by assuming that the auditory system takes a number of short “looks” at the signal, and that these looks are combined in the manner described by signal detection theory (Viemeister and Wakefield, 1991; Dai and Wright, 1995).

Another attempt to use compression to model temporal integration, this time for suprathreshold stimuli, comes from Zwicker’s (1977) “loudness meter,” in which the compressive nonlinearity occurs before long-time-constant temporal integration, and is preceded only by a critical-band filter and a low-pass filter with a time constant of 2 ms. Unfortunately, an evaluation of this procedure is hampered by the considerable variation across studies in the amount of temporal integration observed, with evidence for simple energy integration (Boone, 1973; Florentine *et al.*, 1996, midlevels), as well as for greater (Algom *et al.*, 1989) and lesser (Florentine *et al.*, 1996, low and high levels) amounts. It should also be noted that, although Zwicker’s loudness meter can account for findings such as the temporal variation in the loudness of modulated tones, it is not clear that this success depends on the order in which compression and temporal integration occur.

Perhaps the best psychophysical evidence for fast-acting compression comes from a study by Oxenham and Moore (1994, 1995), who have recently succeeded in modeling excess masking by applying a compressive nonlinearity to each auditory-filter output, followed by smoothing with a temporal window and the addition of a fixed-level internal noise. Because the compression was applied to the filter output on a point-by-point basis, they were able to account for data obtained with very brief (2 ms) signals separated from the maskers by very short silent intervals. This suggests that, although they did not explicitly consider amplitude fluctuations in the filter response to a single masker, an approach similar to theirs could be used to account for our data. The major modification which needs to be made is, as discussed above, the replacement of the power-law nonlinearity with a function which is less compressive at low than at medium levels; such a modification has recently been successfully incorporated into a model of forward and backward masking (Oxenham and Moore, 1997). In addition, it is worth remembering that the gammatone filter should not be used as a front end to such models, as its responses to the positive- and negative-phase maskers used here do not differ markedly in peakiness (Kohlrausch and Sander, 1995). Modifications both to the phase and amplitude characteristics of the gam-

matone filterbank are suggested elsewhere (Glasberg and Moore, 1990; Rosen and Baker, 1994; Kohlrausch and Sander, 1995; Irino and Patterson, 1997; Carlyon and Datta, 1997).

3. Loudness of waveforms differing in crest factor

Preece and Wilson (1988) generated three complex tones with identical power spectra (equal-amplitude components at 1020, 1040, 1060, 1080, and 1100 Hz) but whose waveforms differed in peakiness, having “crest factors” (peak amplitude/rms: ANSI, 1977) of 1.8, 2.6, and 3.2. They reported that listeners judged pairs of these complexes with equal rms amplitudes to have approximately equal loudnesses. This is roughly consistent with the results of experiment 4, which showed that the difference in loudness between the negative- and positive-phase complexes was reduced or reversed when only five components were presented. It seems that, in order for two complexes with identical power spectra to differ markedly in loudness, the outputs of at least some auditory filters must show the marked difference in peakiness which can only arise when a large number of components interact within their passbands.

B. Modeling the compressive nonlinearity at high levels

Many previous attempts to model auditory compression have used a power function, which, when applied to the amplitude (rather than to the power) of the filtered waveform, has an exponent which varies across listeners and studies but which is generally about 0.4 to 0.5 (Zwicker, 1958; Humes and Jesteadt, 1989, 1991; Oxenham and Moore, 1994). A number of authors (e.g., Oxenham and Moore, 1994) have pointed out that this function is similar to that observed on the BM at medium-to-high levels, and have suggested that their data may reflect this compressive nonlinearity. In this subsection we consider whether a similar function can account for the difference in forward masking produced by the high-level positive- and negative-phase maskers of experiment 1.⁴

An ideal approach would be to consider the outputs of an 1100-Hz auditory filter to the positive- and negative-phase maskers of experiment 2, and determine the compression which would mimic the 10- to 15-dB difference in masking effectiveness observed for most listeners at high levels (Fig. 4). Unfortunately, although the MPPs for these two maskers show that the corresponding filter outputs differ markedly in peakiness, they do not provide sufficient information to determine the exact form of those outputs. Our solution was to generate two filter outputs which differed maximally in peakiness, the only limitation on this difference arising from the limited number of components passed by the auditory filter. We then applied a power-law compression to the outputs, and, for a range of exponents, measured the additional attenuation which had to be applied to the less-peaked output in order to equate the resulting measures of excitation. This then provided an estimate of the minimum amount of compression that must be applied to the stimuli. If, as seems probable, our simulated filter outputs differed in

TABLE II. The first two columns show the frequencies and gains of the components used in the filter output simulation discussed in Sec. VI B. The “minimally peaked” output was generated using the phases shown in column three; for the “maximally peaked” output, all components were set to cosine phase.

| Freq. (Hz) | Gain (dB) | Phase (rad) |
|------------|-----------|-------------|
| 200.00 | −27.23 | 0.00 |
| 300.00 | −24.61 | 6.13 |
| 400.00 | −22.02 | 5.76 |
| 500.00 | −19.44 | 5.45 |
| 600.00 | −16.80 | 5.46 |
| 700.00 | −13.97 | 5.70 |
| 800.00 | −10.68 | 6.11 |
| 900.00 | −6.80 | 0.52 |
| 1000.00 | −2.66 | 1.80 |
| 1100.00 | 0.00 | 4.20 |
| 1200.00 | −3.22 | 2.45 |
| 1300.00 | −8.71 | 1.76 |
| 1400.00 | −14.90 | 1.53 |
| 1500.00 | −21.45 | 1.45 |
| 1600.00 | −28.22 | 1.40 |
| 1700.00 | −35.13 | 1.49 |
| 1800.00 | −42.13 | 1.78 |
| 1900.00 | −49.22 | 2.33 |
| 2000.00 | −56.36 | 2.23 |

peakiness by more than did the actual responses to the positive- and negative-phase maskers, then we will have underestimated the compression necessary to account for the data.

The 2nd–20th harmonics of a 100-Hz F_0 were set to amplitudes corresponding to the gain of an auditory filter⁵ centered on 1100 Hz, as shown in the first two columns of Table II. The maximally peaked waveform was generated by summing all of the scaled components in cosine phase, and is shown in Fig. 1(a). The minimally peaked waveform [Fig. 1(b)] was obtained by applying an algorithm (Pumplin, 1985) which minimizes the deviation in the instantaneous power of a waveform from its average value. Each simulated filter output was then full-wave rectified, and excitation (E) was calculated from resulting function $x(n)$ according to the following formula:

$$E = \frac{20 \log \frac{\sum_{n=1}^N x(n)^p}{N}}{p}, \quad (1)$$

where N is the number of samples (2000) in the 10-ms period of each waveform, and p is the exponent of the power-law compressor. The quantity E has the property that, for two waveforms $x(n)$ and $y(n)$, when $E_{x(n)} - E_{y(n)} = z$, attenuating $x(n)$ by z dB will cause the two values of E to be equal. This means that a difference between the values of E produced by two maskers, for example the positive- and negative-phase complexes of experiment 2, can be directly compared to the differences in their physical levels necessary to mask a given signal.

Table III shows the difference between the values of E produced by the minimally peaked and maximally peaked waveforms, for values of the exponent p ranging from 0.1 to 1.0. Recall that in experiment 2 [Fig. 3(b)], at high overall levels, the positive-phase complex needed to be set to a level 10–15 dB higher than that of the negative-phase complex in

TABLE III. Difference between the values of E in Eq. (1), produced by the minimally peaked and maximally peaked waveforms described in Sec. VI B, for values of the exponent p ranging from 0.1 to 1.0.

| p | ΔE |
|-----|------------|
| 0.1 | 10.4 |
| 0.2 | 9.5 |
| 0.3 | 8.7 |
| 0.4 | 8.0 |
| 0.5 | 7.2 |
| 0.6 | 6.5 |
| 0.7 | 5.8 |
| 0.8 | 5.2 |
| 0.9 | 4.6 |
| 1.0 | 3.2 |

order to mask the signal. (For one out of the five listeners this difference was as high as 26 dB.) The predicted differences shown in Table III are generally smaller than these values: Even for an exponent as low as 0.1, the predicted difference is only 10.3 dB, which falls at the lower end of the differences observed experimentally. Thus our approach to modeling the data obtained here at high levels requires a compression more severe than that typically used elsewhere to model excess masking or the growth of loudness. The next section discusses the possible contribution to this compression of different stages of auditory processing.

C. Compression in higher stages of the auditory system

We have argued that the experiments described here have revealed the existence of a compressive nonlinearity which is fast acting and which, compared to a power function, is more compressive at high levels and less compressive at low levels. Although it seems highly likely that the results were influenced by the level-dependent fast-acting nonlinearities observed in the responses of the BM and IHCs (Russell and Sellick, 1978; Sellick *et al.*, 1982; Ruggero *et al.*, 1992), it is worthwhile considering the possibility that compression in neural sites may have had some additional effect. Particularly relevant are the data of Horst *et al.* (1990), who measured the responses of auditory-nerve (AN) fibers to groups of equal-amplitude harmonics that had been summed in cosine phase. They pointed out that the valleys in the waveform of such stimuli have a different spectrum from that of the peaks, and showed that, at high levels, the representation of this part of the spectrum was enhanced in the responses of single AN fibers, as measured by the FFT of period histograms. Because this effect was much smaller at low input levels, their finding reflects a compression which is level dependent. Furthermore, the fact that the peaks and valleys were differentially affected implies that the compression must, like that observed here, have a time constant which is shorter than the period of the waveform. However, it should be borne in mind that, as Horst *et al.* pointed out, it is difficult to distinguish any potential neural basis for compression from the possibility that the measurements obtained in the AN simply reflect the compression already observed in the cochlea.⁶ Nevertheless, the fact remains that the results described here may be influenced by compression at all

stages of the auditory system prior to that at which detection (experiments 1–3) or loudness judgments (experiment 4) occur, and that the number and nature of compressive stages involved may differ across psychophysical tasks.

A related point concerns the implications for the interpretation of experiments 1–3 and of the analysis in Sec. VI B of the nature of the auditory processes underlying forward masking, which has been attributed variously to perseverance of masker excitation (Zwicker, 1977), to smoothing by a centrally located “temporal window” (Moore *et al.*, 1988; Plack and Moore, 1990), and to adaptation (Nelson and Freyman, 1987; Dau *et al.*, 1996a, 1996b). Specifically, if post-stimulus adaptation at some neural site increases as a compressive function of instantaneous firing rate, then this should influence our forward-masking data, but only to the extent that forward masking is mediated by adaptation at that site.⁷ It is worth noting that, even if post-stimulus adaptation did influence the results of experiments 1–3, it would have had no effect on loudness judgements of experiment 4.

The above discussion suggests that one should not be too surprised when a given psychophysical measure of compression does not agree exactly with that observed on the BM, an observation which also applies to previous psychophysical measures of compression. Hence, in order to obtain a complete understanding of the various sources of compression influencing auditory perception, it may be necessary to consider data obtained using a range of techniques. The novel constraint imposed by the current results is that they reflect a compression which must, in addition to being level dependent, act sufficiently quickly to follow the fast fluctuations in the BM response to steady-state complex tones.

ACKNOWLEDGMENTS

We thank Professor Jon Pumplin for providing a copy of his computer program. Brian Moore, Chris Plack, Armin Kohlrausch, and an anonymous reviewer contributed many helpful comments on a previous version of this article.

¹There is a choice of quantitative measures which describe the extent to which the power of a signal fluctuates over time. Three of these, the crest factor, the fourth moment of the waveform, and the fourth moment of the envelope, are described and compared by Hartmann and Pumplin (1988). The three measures generally correlate well with each other. The present article discusses waveforms which differ markedly from each other on all three of the measures, and, for this reason, the general term “peakiness” is used throughout.

²One further listener (PM) initially showed no consistent tendency to judge either complex as louder. He was retested after 2 hours’ practice on other loudness judgment tasks (involving comparisons of two sinusoidally frequency modulated pure tones), at which time he judged the negative-phase components to be louder on 78% of occasions. Unfortunately, he was not retested on the other two conditions (“central components alone” and “off-frequency” components), and so his data are not shown in the figure.

³Oxenham and Moore (1995) state that, if the output of their linear temporal integrator in response to a single masker (M_1) and signal (S) is $M_1^q + S^p$, then the response to two equally effective maskers will equal $2M_1^q + S^p$ (where p and q are the exponents characterizing the compression applied to the signal and masker, respectively). Thus in order for the signal to remain at threshold, S^p must double, which means that the signal level must be increased by $2^{1/p}$, irrespective of the compression (q) applied to the maskers. A similar argument applies whatever form of compressive nonlinearity is applied to the stimuli. Interestingly, it does not apply to our experiment 3b, where the levels of two maskers having different durations

were adjusted so as to just mask the signal. Assuming, for simplicity, a rectangular temporal integrator, the effect of a masker having duration T and level M is $T \cdot F(M)$ and that of the signal is $G(S)$, where F and G are compressive functions which may or may not be identical. When T is reduced then the amount by which M must be increased for the masker to maintain the same effect depends only on the inverse of $F(M)$.

⁴In a companion article (Carlyon and Datta, 1997) we use a cochlear model (Giguère and Woodland, 1994) to simulate the MPPs obtained with positive- and negative-phase maskers. Preliminary attempts to use the model to simulate forward-masked thresholds were unsuccessful: The simulations predicted only very small amounts of forward masking, which, in contrast to the results of experiment 1, depended critically on the phase at which the masker was turned off. One interpretation of this is that forward masking reflects a more long-term measure of masker excitation than is captured by the model.

⁵There is some debate concerning the amplitude response of auditory filters at high levels (Glasberg and Moore, 1990; Rosen and Baker, 1994). We used a filter shape obtained using the “POLYFIT” procedure described by Rosen and Baker (1994), and derived from the data presented by Baker *et al.* (1994) for a 70-dB 1-kHz signal. The fit was based on the roex (p, w, t) filtershape originally described by Patterson *et al.* (1982), with the three parameters set to 18.77, -8.45 , and 7.72 on the low-frequency side and to 19.31 , -33.01 , and 25.4 on the high-frequency side. The resulting filter was somewhat broader than that proposed by Glasberg and Moore (1990). One reason for choosing the broader filter was to generate filter outputs which differed as much as possible in peakiness so as to be sure of not overestimating the amount of compression needed to account for the data.

⁶Horst *et al.* (1990) reported that the maximum instantaneous firing rates in response to some cosine-phase complexes reached a maximum at levels below that at which the average rate saturated. They argued that, at this maximum in instantaneous firing rate, AN fibers were firing on every cycle of the waveform that corresponded to a peak, and that the maximum corresponded to a theoretical limit on instantaneous rate. This would imply that the AN applies a saturating nonlinearity over and above that observed in the responses of the BM and IHCs. However, Horst *et al.* (1990) also pointed out that, at levels above that producing the maximum in instantaneous rate, the average rate continued to increase due to the response to valleys in the waveform “stealing spikes from larger peaks” (p. 2663). Because our results do not depend on the instantaneous representation of the signal, but rather on some long-term measure more akin to the mean firing rate, they will presumably be affected by the limitation in instantaneous rate only to the extent that it modifies the mean rate. Given that Horst *et al.*’s data show that the mean rate can increase while the maximum instantaneous rate is decreasing, it is not clear that this modification is substantial.

⁷Strictly speaking, one should describe such compression as applying not to auditory excitation, but to an aftereffect of that excitation.

Algom, D., Rubin, A., and Cohen-Raz, L. (1989). “Binaural and temporal integration of the loudness of tones and noises,” *Percept. Psychophys.* **46**, 155–166.

ANSI (1969). ANSI S3.6-1969, “Specifications for audiometers” (American National Standards Institute, New York).

ANSI (1977). ANSI S100-1977, “IEEE Standard dictionary of electrical and electronic terms” (American National Standards Institute, New York), 2nd ed.

Baker, R. J., Darling, A., and Rosen, S. (1994). “Human auditory filter nonlinearity across frequency,” University College of London, Speech, Hear. Lang.: Work Prog. **8**.

Boone, M. M. (1973). “Loudness measurements on pure tone and broad band impulsive sounds,” *Acustica* **29**, 198–204.

Carlyon, R. P. (1988). “The development and decline of forward masking,” *Hearing Res.* **32**, 65–80.

Carlyon, R. P. (1996). “Spread of excitation produced by maskers with damped and ramped envelopes,” *J. Acoust. Soc. Am.* **99**, 3647–3655.

Carlyon, R. P., and Datta, A. J. (1997). “Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components,” *J. Acoust. Soc. Am.* **101**, 3648–3657.

Dai, H., and Wright, B. A. (1995). “Detecting signals of unexpected or uncertain durations,” *J. Acoust. Soc. Am.* **98**, 798–806.

Dau, T., Püschel, D., and Kohlrausch, A. (1996a). “A quantitative model of

- the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.
- Dau, T., Püschel, D., and Kohlrausch, A. (1996b). "A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.
- Florentine, M., Buus, S., and Poulsen, T. (1996). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.
- Giguère, C., and Woodland, P. C. (1994). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," J. Acoust. Soc. Am. **95**, 331–342.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," Hearing Res. **47**, 103–138.
- Hartmann, W. M., and Pumpin, J. (1988). "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.
- Hellman, R. P., and Zwislöck, J. J. (1963). "Monaural loudness function at 1000 cps and interaural summation," J. Acoust. Soc. Am. **35**, 856–865.
- Horst, J. W., Javel, E., and Farley, G. R. (1990). "Coding of spectral fine structure in the auditory nerve. II. Level-dependent nonlinear responses," J. Acoust. Soc. Am. **88**, 2656–2681.
- Humes, L. E., and Jesteadt, W. (1989). "Models of the additivity of masking," J. Acoust. Soc. Am. **85**, 1285–1294.
- Humes, L. E., and Jesteadt, W. (1991). "Models of the effects of threshold on loudness growth and summation," J. Acoust. Soc. Am. **90**, 1933–1943.
- Irino, T., and Patterson, R. D. (1997). "A Time-domain, level-dependent auditory Filter: The gammachirp," J. Acoust. Soc. Am. **101**, 412–419.
- Jesteadt, W., Bacon, S. P., and Lehman, J. R. (1982). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.
- Kidd, G., and Feth, L. L. (1982). "Effects of masker duration in pure-tone forward masking," J. Acoust. Soc. Am. **72**, 1384–1386.
- Kidd, G., Mason, C. R., and Feth, L. L. (1984). "Temporal integration of forward masking in listeners having sensorineural hearing loss," J. Acoust. Soc. Am. **75**, 937–944.
- Kohlrausch, A., and Sander, A. (1995). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.
- Levitt, H. (1971). "Transformed up-down methods in psychophysics," J. Acoust. Soc. Am. **49**, 467–477.
- Meddis, R. (1988). "Simulation of auditory-neural transduction: Further studies," J. Acoust. Soc. Am. **83**, 1056–1063.
- Meddis, R., and Hewitt, M. (1991). "Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.
- Moore, B. C. J. (1978). "Psychophysical tuning curves measured in simultaneous and forward masking," J. Acoust. Soc. Am. **63**, 524–532.
- Moore, B. C. J., and Glasberg, B. R. (1981). "Auditory filter shapes derived in simultaneous and forward masking," J. Acoust. Soc. Am. **69**, 1003–1014.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
- Moore, B. C. J., and Glasberg, B. R. (1996). "A revision of Zwicker's loudness model," Acustica united with Acta Acustica **82**, 335–345.
- Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (1988). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102–1116.
- 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.
- Nelson, D. A., and Freyman, R. L. (1987). "Temporal resolution in sensorineural hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 709–720.
- Oxenham, A. J., and Moore, B. C. J. (1994). "Modeling the additivity of nonsimultaneous masking," Hearing Res. **80**, 105–118.
- Oxenham, A. J., and Moore, B. C. J. (1995). "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. (1997). "Modeling the effects of peripheral nonlinearity in listeners with normal and impaired hearing," in *Modeling Sensorineural Hearing Loss* edited by W. Jesteadt (Erlbaum, Hillsdale, NJ) (in press).
- Patterson, R. D. (1994). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. **96**, 1409–1418.
- 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.
- Penner, M. J. (1979). "Forward masking with equal-energy maskers," J. Acoust. Soc. Am. **66**, 1719–1724.
- Penner, M. J. (1980). "The coding of intensity and the interaction of forward and backward masking," J. Acoust. Soc. Am. **67**, 608–616.
- Penner, M. J., and Shiffrin, R. M. (1980). "Nonlinearities in the coding of intensity within the context of a temporal summation model," J. Acoust. Soc. Am. **67**, 617–627.
- Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (1996). "Effects of level and frequency on the detection of decrements and increments in sinusoids," J. Acoust. Soc. Am. **97**, 3791–3799.
- Plack, C. J., and Moore, B. C. J. (1990). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187.
- Preece, J. P., and Wilson, R. H. (1988). "Detection, loudness, and discrimination of five-component tonal complexes differing in crest factor," J. Acoust. Soc. Am. **84**, 166–171.
- Pumpin, J. (1985). "Low-noise noise," J. Acoust. Soc. Am. **78**, 100–104.
- Robinson, C. E., and Pollack, I. (1973). "Interaction between forward and backward masking: A measure of the integrating period of the auditory system," J. Acoust. Soc. Am. **53**, 1313–1316.
- Rosen, S., and Baker, R. J. (1994). "Characterising auditory filter nonlinearity," Hearing Res. **73**, 231–243.
- Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (1992). "Basilar membrane responses to two-tone and broadband stimuli," Philos. Trans. R. Soc. London, Ser. B **336**, 307–315; also in *Processing of Complex Sounds by the Auditory System*, edited by R. P. Carlyon, C. J. Darwin, and I. J. Russell (Oxford U.P., Oxford, 1992).
- Russell, I. J., and Sellick, P. M. (1978). "Intracellular studies of hair cells in the mammalian cochlea," J. Physiol. (London) **284**, 261–280.
- Schroeder, M. R. (1970). "Synthesis of low peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **IT-16**, 85–89.
- Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," J. Acoust. Soc. Am. **72**, 131–141.
- van der Heijden, M., and Kohlrausch, A. (1995). "The role of envelope fluctuations in spectral masking," J. Acoust. Soc. Am. **97**, 1800–1807.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
- Zwicker, E. (1958). "Über psychologische und methodische Grundlagen der Lautheit," Acustica **8**, 237–258.
- Zwicker, E. (1977). "Procedure for calculating loudness of temporally variable sounds," J. Acoust. Soc. Am. **62**, 675–682.
- Zwicker, E., and Scharf, B. (1965). "A model of loudness summation," Psychol. Rev. **72**, 3–26.
- Zwislöck, J. J. (1965). "Analysis of some auditory characteristics," in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York), pp. 1–97.
- Zwislöck, J. J., Piroda, E., and Rubin, H. (1959). "On some poststimulatory effects at the threshold of audibility," J. Acoust. Soc. Am. **31**, 9–14.

Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components

Robert P. Carlyon and A. Jaysurya Datta

MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

(Received 11 June 1996; revised 1 December 1996; accepted 4 March 1997)

Masking period patterns (MPPs) were obtained for maskers consisting of harmonics 2–20 of a 100-Hz fundamental. The signal was always a 5-ms 1100-Hz sinusoid presented 152, 154, 156, 158, or 160 ms after the start of a 400-ms masker. Experiment 1 replicated the finding that, for a masker level of 69 dB component, the shape of the MPP depended strongly on the phases of the components: Summing them in positive Schroeder phase led to a threshold variation of about 18 dB across the MPP, but summing them in negative Schroeder phase produced a flat MPP [A. Kohlrausch and A. Sander, *J. Acoust. Soc. Am.* **97**, 1817–1829 (1995)]. Reducing the level of the positive-phase masker resulted in a systematic flattening of the MPP, whereas the negative-phase MPPs were flat both at high and at low levels. Experiment 2 showed that removing all components of a positive-phase masker except those close to the signal raised thresholds at the minimum of the MPP. In contrast, a similar manipulation applied to the negative-phase masker produced a uniform elevation of the MPP. Experiment 3 showed that an analogous effect could be obtained by manipulating the phases of masker components remote from the signal. It is shown that several features of the data can be simulated using a nonlinear model of the auditory periphery [C. Giguère and P. C. Woodland, *J. Acoust. Soc. Am.* **95**, 331–342 (1994)]. © 1997 Acoustical Society of America. [S0001-4966(97)06306-6]

PACS numbers: 43.66.Dc, 43.66.Mk [WJ]

INTRODUCTION

There exists a large amount of data describing the frequency selectivity of the auditory system, and, in particular, of the amplitude characteristics of the bank of bandpass filters used to model this selectivity. Not only is there qualitative agreement between measures obtained using different techniques (Fletcher, 1940; Zwicker, 1952; Zwicker *et al.*, 1957; Patterson, 1976), but quantitative descriptions of the filter shapes exist for a wide range of center frequencies (CFs) and levels (Patterson *et al.*, 1982; Lutfi and Patterson, 1984; Rosen and Baker, 1994). In contrast, rather little is known about the corresponding phase characteristics.

Recently, Kohlrausch and Sander (1995) described a novel set of experiments, the results of which imposed some constraints on the phase response of the auditory filterbank. One of their experiments measured the threshold for a brief 1100-Hz sinusoid as a function of its temporal position in a harmonic complex masker. This function, termed the “masking period pattern (MPP),” was obtained for three different maskers, each of which had the same power spectrum, consisting of harmonics 2–20 of a 100-Hz fundamental. When the masker harmonics were summed in sine phase, producing a highly peaked waveform, the MPP was also modulated, with the signal threshold varying by about 17 dB as it was stepped through one period of the masker. In contrast, summing the components in “negative” Schroeder phase ($\phi_n = -\pi n(n-1)/N$) resulted in a waveform with a flat envelope (Schroeder, 1970), and a correspondingly flat MPP. The surprising result was that summing the components in “positive” Schroeder phase [$\phi_n = \pi n(n-1)/N$], which also pro-

duced a waveform with a flat envelope, resulted in a signal threshold which varied by as much as 17 dB across the MPP. Kohlrausch and Sander (1995) concluded that the phase response of the 1100-Hz auditory filter transformed the flat envelope of the input into an output containing marked peaks and dips, and that this led to the substantial threshold variations seen in the MPP.

Our interest in Kohlrausch and Sander’s paradigm was originally sparked by the observation, described in a companion article (Carlyon and Datta, 1997), that their stimuli could be used to study the effects of compression in the auditory system. This led to a series of experiments investigating the effects of level and of components remote from the signal frequency on MPPs for positive- and negative-phase maskers. The results of these experiments, presented here, show that the MPP for a positive-phase masker changes shape with level, and is affected by components more than half an octave away from the signal frequency. In the final section we show that it is possible to simulate several aspects of the data using a nonlinear model of the auditory periphery, as described by Giguère and Woodland (1994).

I. EXPERIMENT 1: EFFECT OF LEVEL

A. Method

MPPs were measured for two maskers, each consisting of equal-amplitude harmonics 2–20 of a 100-Hz fundamental. The phase of the n th harmonic was equal to either $\pi n(n-1)/19$ (“positive phase”) or $-\pi n(n-1)/19$ (“negative phase”). Both of these phase relationships produce very flat envelopes; inspection of the stimuli used here

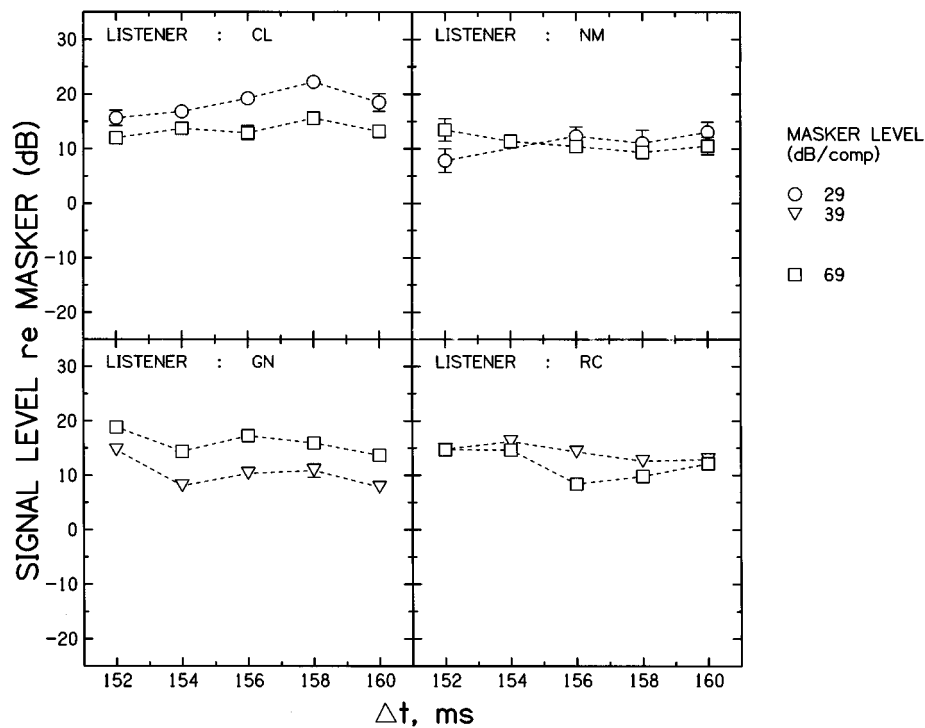


FIG. 1. MPPs obtained with the negative-phase maskers of experiment 1. The level of each masker component was either 29 dB (circles), 39 dB (inverted triangles), or 69 dB (squares). Error bars show plus and minus one standard error.

reveals that they are simply time-reversed and inverted versions of each other. For three listeners, the positive-phase MPP was measured at masker levels of 69, 59, 49, 39, and 29 dB SPL/component; the MPP was also measured for a fourth listener at levels of 69 and 39 dB/component. The negative-phase MPP was measured at two levels: 69 dB and a low level which was 29 dB for two of the four listeners and 39 dB for the other two.¹ MPPs were not measured at intermediate levels of the negative-phase maskers because pilot experiments revealed that those at the two extreme levels were very similar. For reference, it is worth noting that our maskers were identical to those in one condition of the study by Kohlrausch and Sander (1995), with the exception that they used a level of approximately 63 dB/component.

In all conditions the masker duration was 400 ms including two 5-ms raised-cosine ramps. The signal was always a 5-ms 1100-Hz sinusoid composed of two 2.5-ms raised-cosine ramps (no steady state), added in phase to the 1100-Hz masker component. It was turned on 152, 154, 156, 158, and 160 ms after the beginning of the masker (zero-voltage points), thereby sampling one 10-ms period of the masker every 2 ms.

All stimuli were generated digitally at a sampling rate of 20 000 Hz and played out with 12-bit resolution using a CED 1401 laboratory interface. The masker and signal were played out of separate DACs, low-pass filtered (Kemo VBF25.01, attenuation rate=135 dB/oct; cutoff=8600 Hz), attenuated (TDT PA4), and mixed into one channel of a headphone amplifier. They were presented to one earpiece of a Sennheiser HD414 headset, whose frequency response varied by about 3 dB between 200 and 2000 Hz (B&K artificial ear type 4153, condenser microphone cartridge type 4134).

The variation consisted of a drop in sensitivity from 200 to 400 Hz, followed by a broad minimum between 400 and 700 Hz, and by a gradual rise from 700 to 2000 Hz. The electrical signal at the input to the headphones was not corrected for this.

Four listeners, each with absolute thresholds within 15 dB of the 1969 ANSI standard, listened to the stimuli in a double-walled sound-attenuating booth. Thresholds were estimated using a two-down one-up adaptive procedure (Levitt, 1971), with feedback being provided at the end of each 2IFC (Two Interval Forced Choice) trial. The signal level was decreased by 2 dB after every two correct answers and increased by the same amount after every incorrect answer, with each change from increasing to decreasing level and vice versa defining a turnpoint. Each run terminated after 16 turnpoints, and the estimate for that run obtained by averaging the signal levels at the last 12 turnpoints. All data presented here were obtained from the mean of at least four runs.

B. Results

Figure 1 shows that the MPPs for the negative-phase masker are essentially flat both at high and low levels. In contrast, the open squares in Fig. 2 show that the MPP for a 69-dB positive-phase masker contains a threshold variation of about 18 dB. These findings are very similar to the results of Kohlrausch and Sander (1995), as is the fact that the maximum and minimum thresholds in the positive-phase MPP occur at masker-signal delays (Δt) of 152 and 158 ms, respectively. However, as masker level decreases, the MPP

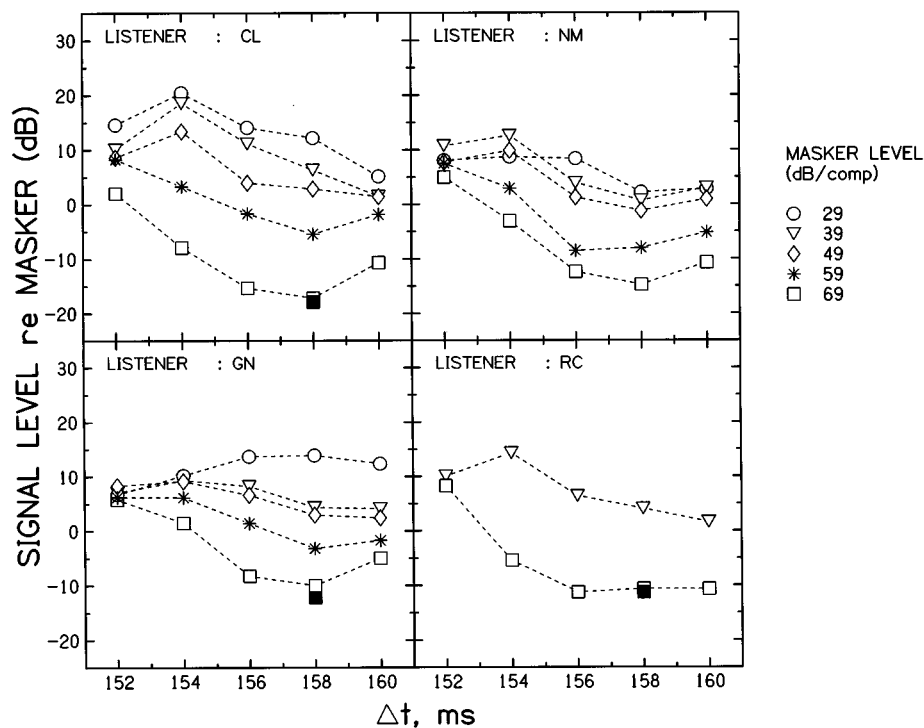


FIG. 2. Open symbols show the MPPs obtained with the positive-phase maskers of experiment 1. The level of each masker component was either 29 dB (circles), 39 dB (inverted triangles), 49 dB (diamonds), 59 dB (asterisks), or 69 dB (squares). The filled squares show the data with a masker level of 69 dB/component in the presence of a 2-kHz low-pass noise (see text for details). Error bars show plus and minus one standard error.

becomes progressively flatter, and there is some evidence of a 2-ms shift of the minimum and maximum of each function towards longer delays.

One possible explanation for the flattening out of the positive-phase MPPs at low levels might be that thresholds in the valleys of the MPP were actually limited by absolute threshold. For a masker level of 39 dB SPL, the lowest masked thresholds were 40 dB SPL for listener CL, 39 dB for NM, 45 dB for GN, and 40 dB for RC. These thresholds were 20, 19, 14, and 14 dB, respectively, above each listener's absolute threshold for the 5-ms signal. These sensation levels suggest that it is unlikely that the flattening out of the MPPs at low levels was due to an absolute threshold limitation. As an extra check, the lowest threshold in the 69-dB positive-phase MPP was remeasured for three listeners in the presence of a 2-kHz low-pass noise. The noise, which had a spectrum level of -6, 8, and 6 dB for listeners CL, GN, and RC, raised their detection thresholds in the absence of a Schroeder-phase masker to 30, 41, and 44 dB, respectively. These levels were 21, 18, and 14 dB below the minimum threshold in the MPP for the 69-dB positive-phase masker, and these amounts were roughly similar to the difference between absolute threshold and the minimum of the 39-dB masker. The filled square in each plot shows that, when the low-pass noise was combined with the 69-dB positive-phase masker, thresholds were similar to those obtained with the positive-phase masker alone.

An alternative explanation for the flattening of MPPs at low levels may be that auditory filters are sharper at low rather than at high levels (Glasberg and Moore, 1990; Rosen and Baker, 1994). Because a highly peaked waveform re-

quires the interaction of a large number of components (Preece and Wilson, 1988, Fig. A1), it may be that the limited bandwidth of these filters precluded a highly peaked output. Evidence relevant to this hypothesis is provided by experiment 2, which examines the effects of removing components with frequencies remote from 1100 Hz on the output of an auditory filter centered on that frequency.

II. EXPERIMENT 2: EFFECT OF THE NUMBER OF COMPONENTS

A. Method

Experiment 2 consisted of three conditions. In the "19-component" condition, MPPs were measured for positive- and negative-phase complexes having a level of 69 dB/component. The stimuli and procedure were as described in Sec. I A. Four listeners took part: two of these had also participated in experiment 1, and their data for this condition were reused here. The second and third conditions involved removing all harmonics except those between 900–1300 Hz inclusive ("five-component condition") or 800–1400 Hz inclusive ("seven-component condition"). In both of these conditions, the phases and amplitudes of those harmonics actually present were the same as in the 19-component condition.

B. Results

The MPPs for the negative-phase maskers of experiment 2 are shown in Fig. 3, whose ordinate spans a smaller range than that of Figs. 1 and 2. The MPPs are generally flat,

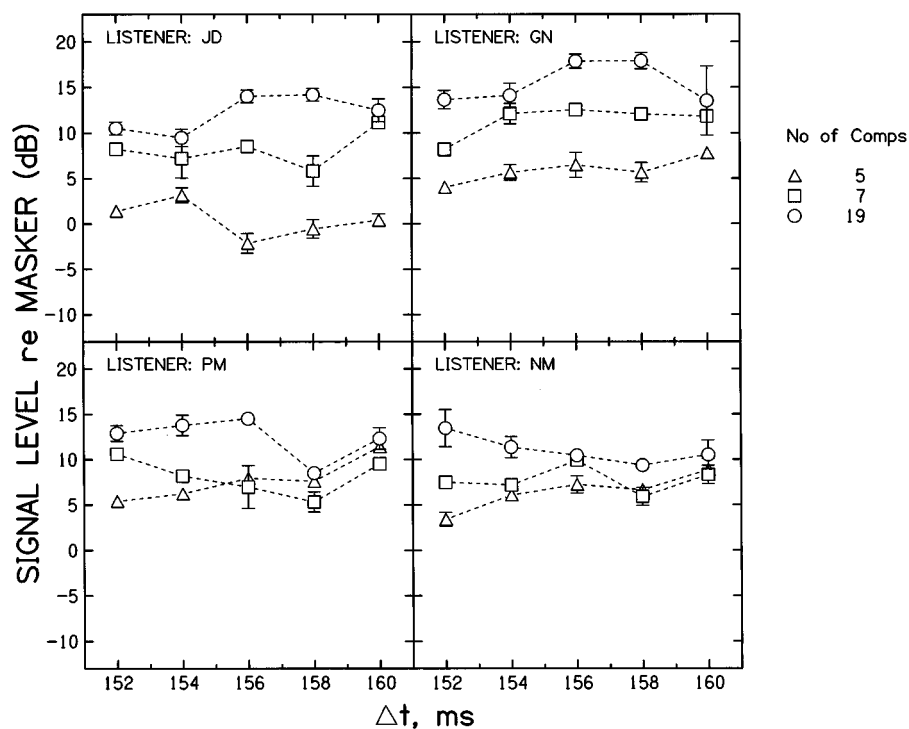


FIG. 3. MPPs for the negative-phase condition of experiment 2. Data are shown for maskers containing five (triangles), seven (squares), and 19 (circles) components. Error bars show plus and minus one standard error.

showing a variation of no more than 5 dB, regardless of the number of components present in the masker. There is a tendency for thresholds to increase as more components are added. This trend may be partly due to the detection of splatter, as the signal spectrum was attenuated *re*: 1100 Hz by

only 5.9 dB at the edges of the five-component complex and by 14.9 dB at the edges of the seven-component complex.

In contrast, Fig. 4 shows that the thresholds in the valleys of the positive-phase MPPs *decrease* by 10–15 dB when the number of components is increased from five to 19. This

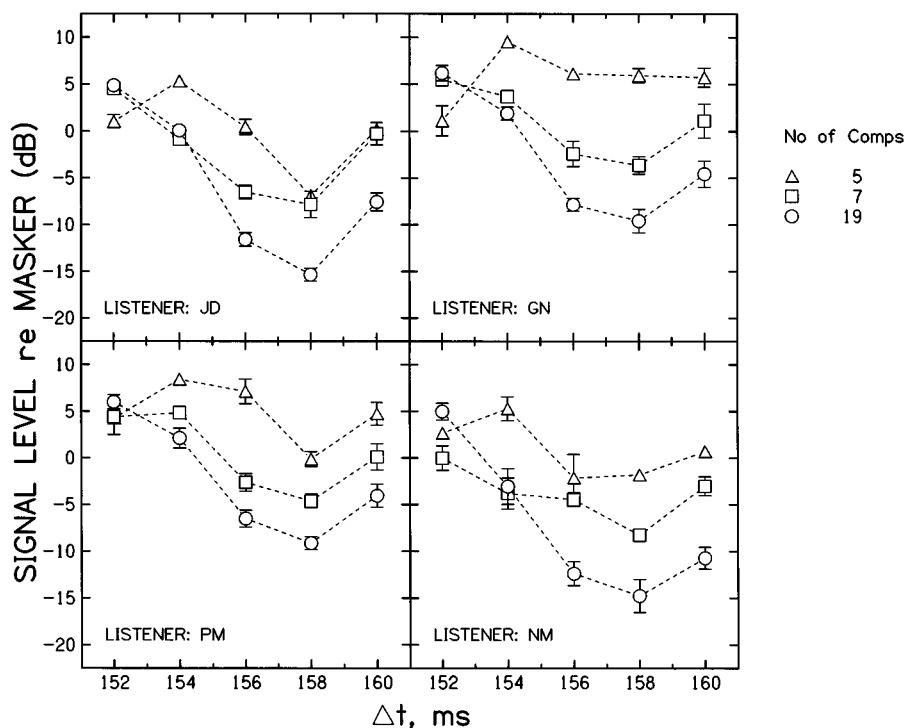


FIG. 4. MPPs for the positive-phase condition of experiment 2. Data are shown for maskers containing five (triangles), seven (squares), and 19 (circles) components. Error bars show plus and minus one standard error.

is consistent with the idea that, as more components are added, the valleys in the filter output become longer and/or reach lower amplitudes. A similar explanation was proposed by Schroeder and Mehrgardt (1982) to explain their finding that the threshold for a long-duration tone dropped as more components were added to a cosine-phase harmonic masker. If the explanation is correct, then the peaks in the filter output should reach higher amplitudes when more components are added, and one might expect a corresponding threshold increase at the maximum of the MPP. However, although the location of the maximum shifts from $\Delta t = 154$ ms with five components to $\Delta t = 152$ ms with 19, there is no evidence that the threshold at the maximum increases. This may be because, although the peak gets larger as more components are added, it also becomes briefer, so that parts of the 5-ms signal actually fall in a valley of the filter output. This “off-time listening” may also explain the finding, apparent in the data of Kohlrausch and Sander (1995) and in our Figs. 1 and 2, that even the highest threshold in the positive-phase MPP is lower than those observed anywhere in the negative-phase MPP. It is worth noting that, if this explanation is correct, then the 18-dB variation in threshold observed in the positive-phase MPP actually underestimates the amplitude variation at the output of the 1100-Hz auditory filter.

III. EXPERIMENT 3: MANIPULATING THE PHASE OF OFF-FREQUENCY COMPONENTS

A. Rationale and method

Experiment 2 showed that the modulation in the output of an 1100-Hz auditory filter is affected by components below 800 Hz and/or above 1400 Hz. Experiment 3a investigated this dependence on off-frequency components in more detail, and used a technique designed to minimize the effects of splatter. Specifically, the masker always contained harmonics 2–20 of the 100-Hz fundamental, and the phases of components remote from 1100 Hz were manipulated. In one condition, the central M components were selected from the positive-phase complex of experiment 1, and the remaining components were selected from the negative-phase complex of that experiment. The opposite was true for the other condition, in which the central M components were in negative phase and the remaining “flanking” components were in positive phase. In both conditions thresholds were obtained for the 5-ms 1100-Hz signal at $\Delta t = 158$ ms (the minimum of the positive-phase MPP) as a function of M . The idea was that, when M is sufficiently large for the off-frequency components not to affect the modulation in the filter output, then further increases in M should not affect thresholds.²

Experiment 3b investigated the relative influence of components below and above 1100 Hz on the output of an auditory filter centered on that frequency. In the “low negative” condition the phases of all components were taken from the positive-phase complex, except for those having frequencies between 200 and 800 Hz inclusive, whose phases were the same as in the negative-phase complex. In the “high negative condition” the only components whose phases were taken from the negative-phase complex were those having frequencies between 1400 and 2000 Hz inclu-

sive; all others were taken from the positive-phase complex. Note that in both conditions the five central (900–1300 Hz) components were in positive phase. As in experiment 3a, the signal was an 1100-Hz 5-ms sinusoid presented at $\Delta t = 158$ ms.

B. Results

The results of experiment 3a are shown by the open symbols of Fig. 5. The condition with the central components in positive phase is shown by the triangles; that with the central components in negative phase is shown by the squares. For all listeners except PM the functions cross at $M = 7$. This means that the threshold with components between 800 and 1400 Hz in positive phase and the rest in negative phase is about the same as when the converse is true. This is striking, because it means that the phase relationship of those components outside the 600-Hz band centered on the signal has as big an effect on threshold as that of those within it. Furthermore, for these same three listeners, threshold is still higher when the 13 harmonics between 500 and 1700 Hz are in positive phase (triangles, $M = 13$) than when all 19 harmonics are in positive phase. This too is striking, because it means that components more than 500-Hz distant from the 1100-Hz signal affect its threshold. The exception to both of these findings is listener PM, whose functions cross at $M = 5$ and for whom thresholds are nearly as low when only the nine central components are in positive phase as when all 19 components share that phase relationship. For him, then, the effect of components remote from 1100 Hz on threshold was smaller than it was for the other listeners.

The relative importance of components above and below 1100 Hz can be compared by inspecting the filled symbols to the left of each panel in Fig. 5. For listeners JD and GN, thresholds are much higher in the “low negative” (circles) than in the “high negative” (diamonds) condition. The data suggest that, for these listeners, the output of an auditory filter centered on 1100 Hz is dominated by components close to and below that frequency. For the other two listeners, however, thresholds in the two conditions are roughly equal, indicating that components above and below 1100 Hz both influence the auditory filter output.

IV. DISCUSSION

A. Comodulation masking release

Experiments 2 and 3 showed that either removing components of the positive-phase complex remote from the 1100 Hz, or manipulating their phases, can produce a marked increase in signal threshold. One possible explanation for this finding comes from the phenomenon of comodulation masking release (“CMR”: Hall *et al.*, 1984), which shows that off-frequency components can reduce thresholds via a central process, provided that they produce modulation in auditory filters remote from the signal frequency (f_s) which is coherent with that in filters tuned to f_s . Thus, if the off-frequency components in our experiments produced such coherent modulation, the effects of removing them might be attributed to a central rather than to a peripheral effect. How-

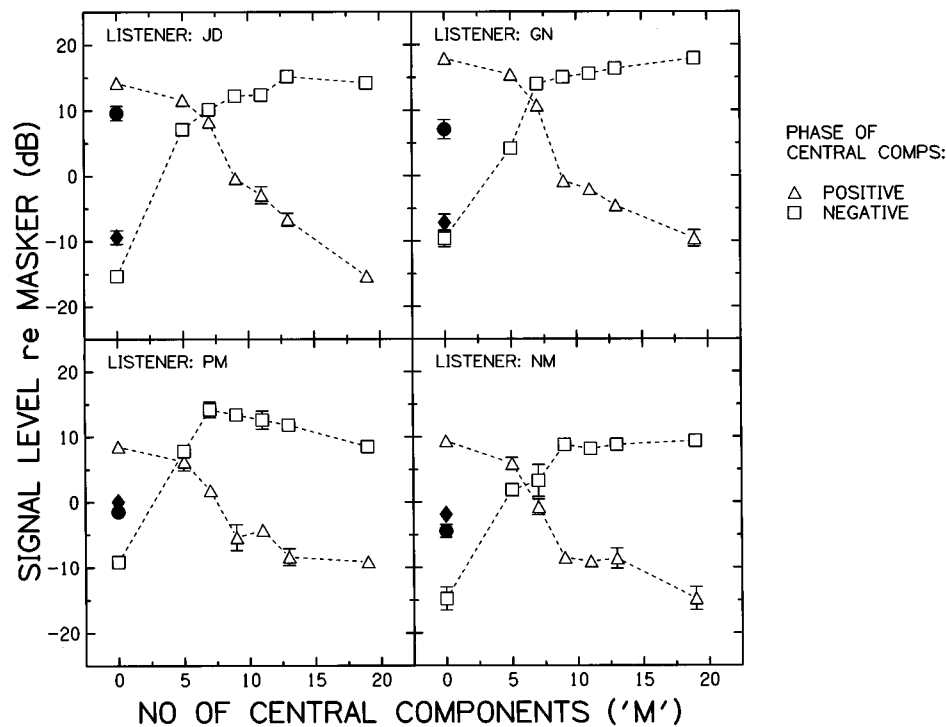


FIG. 5. Results of experiment 3. Open symbols show the results obtained with the central M components in either positive (triangles) or negative (squares) phase. The unconnected filled symbols to the left of each plot are for the "low negative" (circles) and "high negative" (diamonds) conditions. Error bars show plus and minus one standard error.

ever, the data of Kohlrausch and Sander (1995) show that the minimum in the MPP shifts by more than 5 ms as the signal frequency increases from 600 to 1600 Hz. Thus it seems unlikely that listeners could have used common patterns of modulation to detect the signal.

B. Physiological model

1. Background and method

Kohlrausch and Sander (1995) have shown that the output of a passive basilar membrane model (Strube, 1985) at high input levels produces a response to the positive-phase complex which is much more peaked than its response to the negative-phase complex. This is in qualitative agreement with the MPPs obtained experimentally with the two types of masker. However, as they point out, the phase characteristics of the basilar membrane (BM) depend on the stimulus level (e.g., Ruggero *et al.*, 1992), and so it is likely that a nonlinear model of the BM would be needed to account for the effects of masker level reported here.

In this section we present the results of a quantitative simulation of experiments 1–3, using the nonlinear model of the auditory periphery described by Giguère and Woodland (1994). The model includes a simulation of the external and middle ears, of the basilar membrane (BM) and cochlear fluids, and of the effects of the outer hair cells on this vibration. The model, as implemented by Patterson *et al.* (1995), also gives the user a choice of simulating the effects of the inner hair cells (Meddis, 1988), of applying a logarithmic compression to the BM vibration, or, as we chose to do, neither. The same parameters and assumptions are used to

simulate the results of all experiments. The only modification of the model is a change in the parameter " Q_n ," which describes the sharpness of BM tuning, from its default value of 2 to the (sharper) value of 8. A similar modification was made by Carlyon (1996) in order to model the excitation patterns of tones having "damped" and "ramped" envelopes. As was pointed out in that article, the default value of Q_n predicts a degree of frequency selectivity which is much coarser than that observed psychophysically.

The details of our simulation are described in the Appendix. Briefly, each data point was estimated by measuring the output of one frequency channel of the model to the masker alone and to masker-plus-signal waveforms at a wide range of signal-to-masker ratios. After smoothing with a low-pass filter, the masker-plus-signal outputs were divided by those to the masker alone, and the maximum value in each resulting function noted. Threshold was defined as the signal level necessary for this maximum value to exceed a criterion value. This criterion was chosen so that the threshold at the maximum in the positive-phase 69-dB/component MPP would be accurately estimated; the same criterion was used for the simulation of all data points. The frequency channel was selected in order to produce the highest model output to the signal-plus-masker, relative to the output to the masker alone, in the positive-phase condition and with a masker level of 69 dB/component. The same channel was used for all simulations described here. Alternative simulations showed that the center frequency of the optimal channel decreased by about 10% with a 40-dB reduction in masker level. Those simulations also showed that selecting a channel with a lower CF would not have significantly

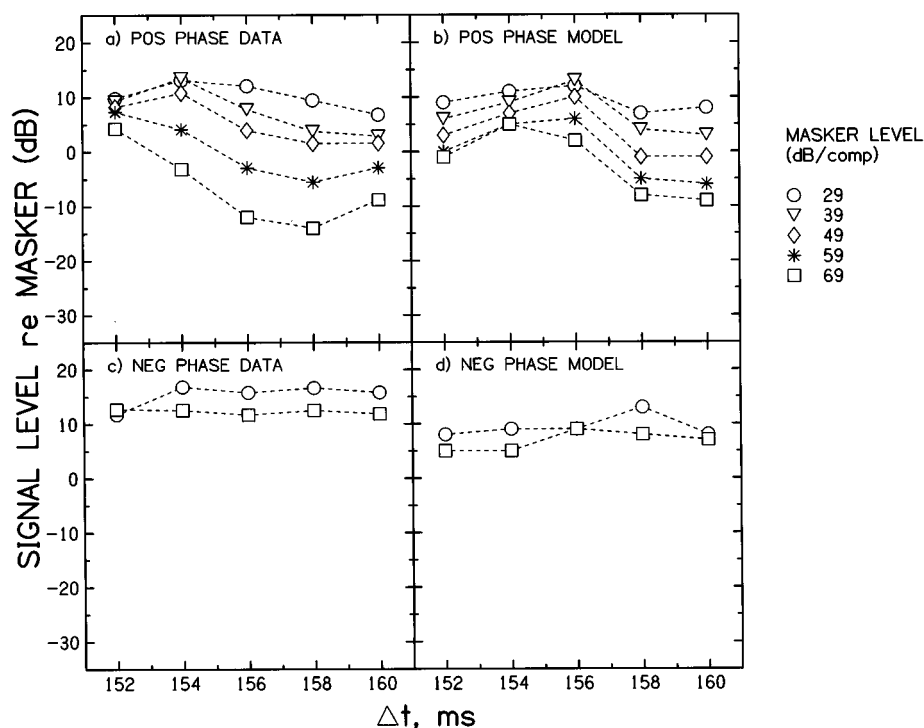


FIG. 6. Panel (a) shows the data from the positive-phase condition of experiment 1 averaged across listeners CL, NM, and GN. Panel (c) shows the corresponding negative-phase data averaged across the two listeners (CL and NM) who were tested using the 29 and 69 dB maskers. Parts (b) and (d) show simulations of parts (a) and (c), respectively.

changed the shapes of the predicted MPPs. It would, however, have slightly reduced the predicted thresholds at low masker levels relative to those at high levels.

2. Results

The results of the simulation of experiment 1 are shown, together with the mean data from that experiment, in Fig. 6. Parts (a) and (b) of the figure show that the simulation captures several aspects of the data obtained with the positive-phase masker. Specifically, it successfully predicts the flattening of the MPP as masker level is reduced, and the fact that this flattening results from an increase in relative thresholds near the minimum of the MPP rather than from a decrease near the maximum. The size of the threshold variation in the positive-phase MPP is slightly underestimated for the highest masker level, being about 18 dB in the data but only 14 dB in the simulation. The major discrepancy between simulation and data is in the location of the MPP maximum, which at the highest level occurs at $\Delta t = 152$ ms in the data but 154 ms in the simulation. The simulation does, however, predict the 2-ms shift in the maximum as masker level is reduced. Also, as Fig. 6(c) and (d) shows, the simulation successfully predicts the flat MPPs for the negative-phase masker both at low and at high levels. In this case, the overall size of the thresholds predicted by the model are slightly lower than those obtained experimentally.

Figure 7 shows the results of the simulation of experiment 2, together with the mean data obtained. The experimental data show a drop in the minimum of the positive-phase MPP as more components are added, with the opposite

effect occurring for the negative-phase MPP. Both of these trends also occur in the simulation, but to a much smaller extent. In particular, the simulation underestimates the decrease in threshold when the number of negative-phase components is reduced. As discussed in Sec. II B, the large decrease observed experimentally may have been partly due to listeners detecting splatter present in the spectrum of the brief signal. This would not have been captured by the single-channel simulation used here.

Finally, the data and simulations of experiment 3 are shown by the open symbols in Fig. 8. The simulation captures the monotonic change in threshold which occurs in both conditions as M (the number of central components) is increased. It also predicts the crossover in the two functions when M is between five and seven, and the flattening off of the negative-phase curve for values of M above nine. It does, however, underestimate the extent of the threshold variation in each condition.

C. Nonphysiological approaches

Kohlrausch and Sander (1995) have pointed out that the different shapes of the MPPs for positive- and negative-phase maskers are inconsistent with the "gammatone" filter characteristics proposed by Patterson *et al.* (1988). Their arguments can be summarized as follows: (i) Both positive- and negative-phase waveforms have flat envelopes because of the curvature in their phase functions. This curvature (second derivative of the phase function) is constant across frequency, and is equal and opposite for the two maskers, being positive for the positive-phase masker and negative for the

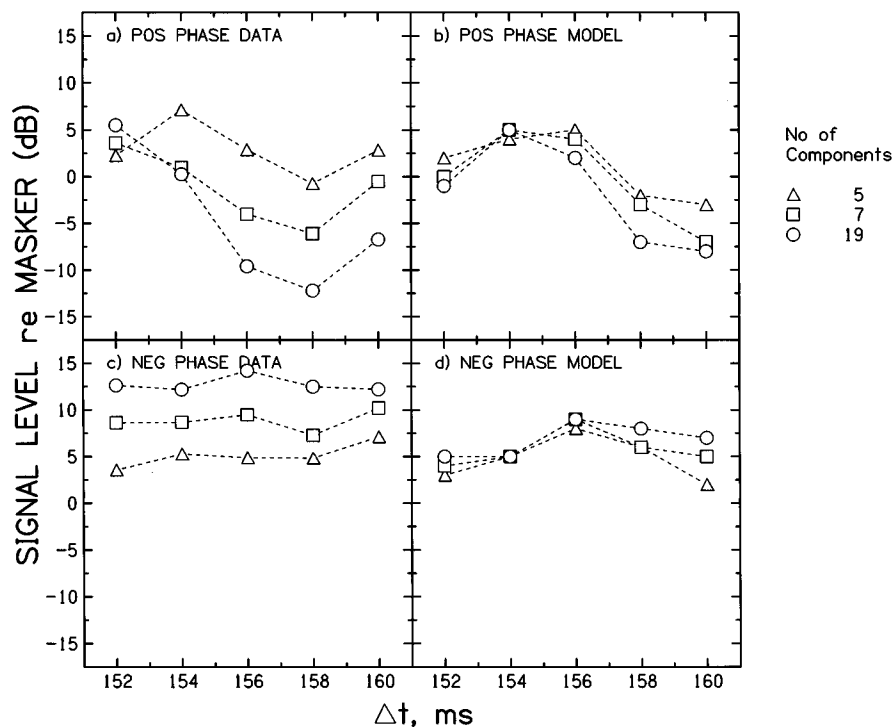


FIG. 7. Mean data and simulations of the results of experiment 2 for maskers containing five (triangles), seven (squares), and 19 (circles) components.

negative-phase masker. (ii) For the auditory filter to produce a peaked output to these stimuli it must reduce this curvature, by having a phase response with equal and opposite curvature to the signal. (iii) Therefore, to account for the positive-phase MPPs, the filter should have negative curvature throughout its passband, but (iv) the phase response of the gammatone is antisymmetric about its CF, having negative curvature below CF and positive curvature above it, which means that (v) the gammatone cannot be used to model the MPPs obtained with Schroeder-phase complexes.

Although it is undoubtedly true that the gammatone cannot account for the results presented here, it may be possible to obtain an approximation to the data without adopting the physiological approach described in the previous subsection. The magnitude response of the auditory filter is asymmetric

at high levels, and, as Baker (personal communication) has pointed out, the direction of this asymmetry means that the peakiness of the filter output will depend mostly on the phase response of the filter below CF. Because this region of the gammatone phase response has negative curvature, then modifying the gammatone to include the asymmetry in the magnitude response may produce peaked outputs in response to the positive-phase complex. Two such modifications, the “all pole gammatone” (Lyon, 1996) and the “gammachirp” filters (Irino and Patterson, 1997) have recently been proposed, the latter of which does indeed produce peakier filter responses to the positive-phase complex than to the negative-phase complex (Irino, personal communication). However, it remains to be seen whether this approach can account for all

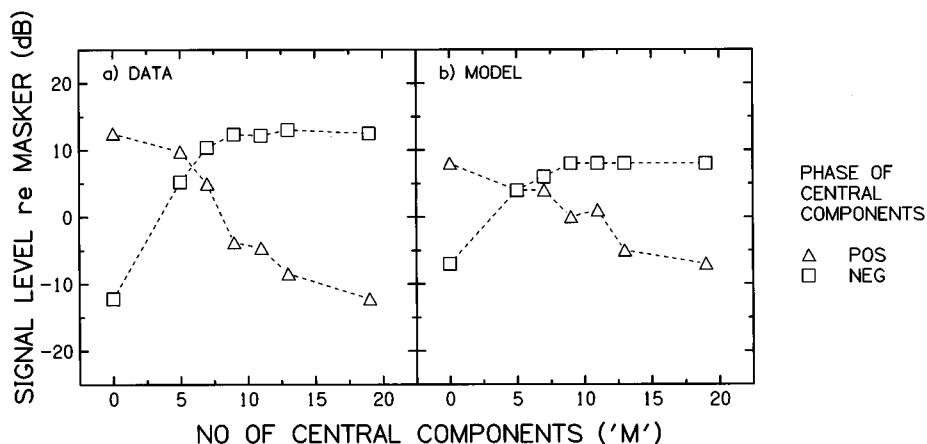


FIG. 8. Open symbols show the mean data [part (a)] and simulation [part (b)] of experiment 3.

of the aspects of the data captured by the simulation presented here.

D. Summary

The experiments described here have extended Kohlrausch and Sander's (1995) finding that the response of an auditory filter centered on 1100 Hz to a positive-phase masker, as measured by MPPs, is much more peaked than that to a negative-phase masker. In addition the data show that the MPPs become progressively flatter as the level of the positive-phase masker is reduced and when components remote from the signal frequency are removed. These two new findings may be related, given that the auditory filter is sharper at moderate levels than at very high levels (Glasberg and Moore, 1990; Rosen and Baker, 1994): Reducing the masker level will reduce the number of components that interact within the passband of the filter, and this may reduce the peakiness of its output. Certainly the fact that, at high levels, MPPs are affected by components more than half an octave distant from f_s imposes constraints not only on the phase characteristics but also on the amplitude response of auditory filters.

ACKNOWLEDGMENTS

We thank Hedwig Gockel, Roy Patterson, Laurent Demany, Armin Kohlrausch, and an anonymous reviewer for helpful comments on a previous version of this manuscript.

APPENDIX: PHYSIOLOGICAL SIMULATION

1. Details of implementation

The simulations shown in Figs. 6–8 were obtained by implementing Giguère and Woodland's (1994) cochlear model using the software platform provided by Patterson *et al.* (1995). To produce inputs to the model, masker waveforms were generated using the same program as for the experiments, and signals added in software at appropriate delays and levels. The simulation consisted of the following stages.

(1) *Channel selection.* The simulation only considers the output of one frequency channel. In order to select this channel, two sounds were input to the model: The positive-phase masker, and the same masker with a signal added at a delay of 158 ms and a signal-to-masker (SMR) ratio of 10 dB. The “input gain” of the model was set to simulate a masker level of 69 dB/component. The output of 11 channels having center frequencies (CFs) from 826.4 to 2143.6 in steps of 10% were analyzed, with the response of each channel to the masker+signal being divided by that to the masker alone. We then selected the channel for which this ratio was greatest at any time between 145 and 185 ms after the onset of the masker. This channel had a CF of 1331 Hz, higher than that of the signal, reflecting the basalward shift of the excitation pattern at high input levels (Johnstone *et al.*, 1986). Alternative simulations revealed that this choice would not have been strongly affected had we used the negative-phase complex and/or other masker-signal delays. It was used for all simulations presented here.

(2) *Find the peak in the positive-phase 69-dB MPP.* Figure 2 shows that the SMR at the peak of this MPP is about 5 dB. Samples of the positive-phase masker mixed with signals at each Δt and with an SMR of 5 dB were input to the 1331-Hz channel of the model, as was the “masker alone” waveform. For each waveform, the channel output was rectified, and then smoothed with a single-stage low-pass filter having a time constant of 3 ms, in order to simulate more central limits on temporal resolution (Viemeister, 1979). The smoothed output for each of the masker+signal waveforms was then divided by that of the masker alone, and portions of the resulting ratio waveform between 145 and 185 ms inspected to determine the maximum change produced by the signal. This quantity, “MAXRAT(Δt),” was defined as the larger of (i) the highest value present in the waveform, and (ii) the reciprocal of the smallest value. In other words, for each Δt , MAXRAT reflected the maximum change produced by the signal, expressed as a proportion of the response to the masker, irrespective of whether this change corresponded to an increase or decrease in output. The Δt at which MAXRAT(Δt) was smallest was defined as the maximum of the MPP, and the value of this MAXRAT(Δt) was termed THRESHRAT. For all remaining simulations it was assumed that a signal was above threshold when MAXRAT(Δt) > THRESHRAT.

(3) *Completing the 69-dB positive-phase MPPs.* For values of Δt which did not correspond to the MPP maximum, masker+signal waveforms were generated for SMRs ranging from –30 to +15 dB in 1-dB steps. The simulation commenced with the highest SMR and was repeated with the SMR decreasing by 1 dB until MAXRAT(Δt) < THRESHRAT. At this point the simulation stopped, and the final SMR was used as the predicted threshold.

(4) *Remaining thresholds.* All remaining simulations were then performed using the procedure described in stage 3 and the value of THRESHRAT obtained in stage 2. Simulation of thresholds at other masker levels could be achieved by varying the input gain of the model. In all other cases a fresh set of masker and masker+signal waveforms were generated for each Δt . For the simulation of experiment 3 a single Δt of 158 ms (the same as in the experiment) was used for all values of M .

2. Choice of parameters

Our implementation of Giguère and Woodland's (1994) model differs from that described in their article in the tuning of local segments of the BM, which we “artificially” sharpened by increasing Q_n from 2 to 8. As Carlyon (1996) has pointed out, the default value of Q_n produces excitation patterns with much shallower upper skirts than those obtained psychophysically: for example, the excitation pattern of an 86 dB SPL 1000-Hz sinusoid is attenuated by only 0.6 dB one octave above its peak. The effect of changing Q_n can be seen in Fig. A1, the solid lines of which show simulations of experiment 3 for values of Q_n ranging from 2–12. The experimental data are replotted in each panel as dotted lines. As shown in panel (a), use of the default parameters leads to the prediction that, in the condition where the central M components are in positive phase (triangles), thresholds should re-

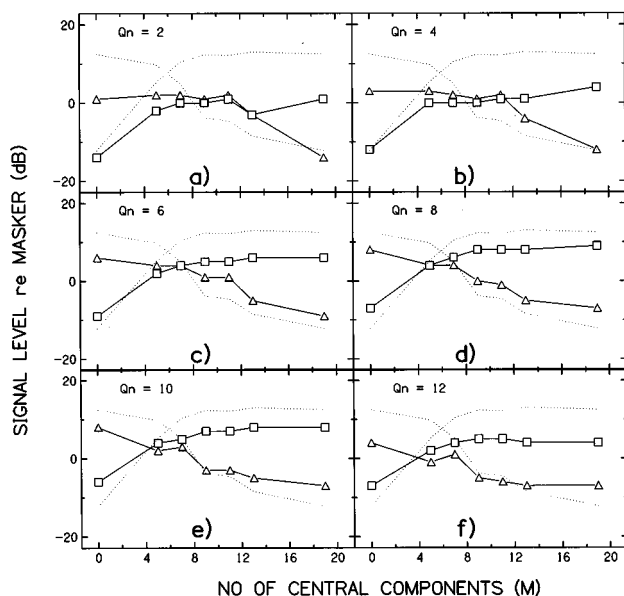


FIG. A1. Solid lines in panels (a)–(f) show simulations of the results of experiment 3, for values of the model parameter Q_n ranging from 2 to 12. In each panel, the mean data from the experiment are replotted as dotted lines.

main constant as M is increased from 0 to 11. This is consistent with the output of the broad 1100-Hz filter in the “default” simulation being *strongly* influenced by negative-phase components remote from 1100 Hz, even when all components within 500 Hz of the center frequency are in positive phase. In contrast, when $Q_n=8$, the simulation matches the data in showing some threshold drop as M is increased beyond 11, but in also exhibiting a systematic decrease as M is increased from 0 to 11.

¹For each listener, one of the levels differed slightly from that stated in the text, both for the positive-phase and negative-phase maskers. Specifically, the stated level of 29 dB was actually 28 dB for listener CL and 23 dB for listener NM, and the stated level of 39 dB was actually 37 dB for listener RC and 41 dB for listener GN. These differences will be ignored in our discussions of the data. They arose because this subset of conditions also formed part of another study (Carlyon and Datta, 1997) which required the masker level to be selected individually for each listener.

²It is worth noting that there is a phase discontinuity at the transition between the central and flanking components. This means that the pairs of components on either side of the transition have a phase relationship different from that defined by either the positive or the negative Schroeder phase formula. Because these formulas differ only by a constant factor of -1 , the size of the discontinuity will be the same when the central components are in positive phase as when they are in negative phase.

ANSI (1969). ANSI S3.6-1969, “Specifications for audiometers” (American National Standards Institute, New York).

Baker, R. (personal communication).

Carlyon, R. P. (1996). “Spread of excitation produced by maskers with damped and ramped envelopes,” *J. Acoust. Soc. Am.* **99**, 3647–3655.

Carlyon, R. P., and Datta, A. J. (1997). “Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,” *J. Acoust. Soc. Am.* **101**, 3636–3647.

Fletcher, H. (1940). “Auditory patterns,” *Rev. Mod. Phys.* **12**, 47–65.

Giguère, C., and Woodland, P. C. (1994). “A computational model of the auditory periphery for speech and hearing research. I. Ascending path,” *J. Acoust. Soc. Am.* **95**, 331–342.

Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notched-noise data,” *Hearing Res.* **47**, 103–138.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). “Detection in noise by spectro-temporal pattern analysis,” *J. Acoust. Soc. Am.* **76**, 50–56.

Irino, T. (personal communication).

Irino, T., and Patterson, R. D. (1997). “A time-domain, level-dependent auditory filter: The gammachirp,” *J. Acoust. Soc. Am.* **101**, 412–419.

Johnstone, B. M., Patuzzi, R. P., and Yates, G. K. (1986). “Basilar membrane measurements and the travelling wave,” *Hearing Res.* **22**, 147–153.

Kohlrausch, A., and Sander, A. (1995). “Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,” *J. Acoust. Soc. Am.* **97**, 1817–1829.

Levitt, H. (1971). “Transformed up-down methods in psychophysics,” *J. Acoust. Soc. Am.* **49**, 467–477.

Lutfi, R. A., and Patterson, R. D. (1984). “On the growth of masking asymmetry with stimulus intensity,” *J. Acoust. Soc. Am.* **76**, 739–745.

Lyon, R. F. (1996). “All-Pole Auditory Filter Models,” in *Diversity in Auditory Mechanics*, edited by E. Lewis (World Scientific, Singapore).

Meddis, R. (1988). “Simulation of auditory-neural transduction: Further studies,” *J. Acoust. Soc. Am.* **83**, 1056–1063.

Patterson, R. D. (1976). “Auditory filter shapes derived with noise stimuli,” *J. Acoust. Soc. Am.* **59**, 640–654.

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., Holdsworth, J., and Rice, P. (1988). “Spiral Vos Final Report, Part A: The Auditory Filterbank,” Applied Psychology Unit, Cambridge, England, Contract Report: APU 2341.

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.

Preece, J. P., and Wilson, R. H. (1988). “Detection, loudness, and discrimination of five-component tonal complexes differing in crest factor,” *J. Acoust. Soc. Am.* **84**, 166–171.

Rosen, S., and Baker, R. J. (1994). “Characterising auditory filter nonlinearity,” *Hearing Res.* **73**, 231–243.

Ruggero, M. A., Rich, N. C., and Recio, A. (1992). “Basilar membrane responses to clicks,” in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 271–278.

Schroeder, M. R. (1970). “Synthesis of low peak-factor signals and binary sequences with low autocorrelation,” *IEEE Trans. Inf. Theory* **IT-16**, 85–89.

Schroeder, M. R., and Mehrgardt, S. (1982). “Auditory masking phenomena in the perception of speech,” in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granström (Elsevier, Amsterdam), pp. 79–87.

Strube, H. W. (1985). “A computationally efficient basilar-membrane model,” *Acustica* **58**, 207–214.

Viemeister, N. F. (1979). “Temporal modulation transfer functions based on modulation thresholds,” *J. Acoust. Soc. Am.* **66**, 1364–1380.

Zwicker, E. (1952). “Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones,” *Acustica* **4**, 415–420.

Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). “Critical bandwidths in loudness summation,” *J. Acoust. Soc. Am.* **29**, 548–557.

Auditory filters measured at neighboring center frequencies^{a)}

Marc A. Fagelson

Department of Communicative Disorders, East Tennessee State University, Johnson City, Tennessee 37614

Craig A. Champlin

Department of Communication Sciences and Disorders and Institute for Neuroscience, University of Texas, Austin, Texas 78712

(Received 12 April 1996; accepted for publication 14 January 1997)

Auditory filters were derived in 20 normal-hearing human listeners at center frequencies (CFs) of 913, 1095, 3651, and 4382 Hz using the roex (p, r) method. Comparisons were made between slopes of the filters' skirts at the neighboring CFs with filter output levels of 45 and 70 dB. The same comparisons were made with regard to filter equivalent rectangular bandwidth (ERB). In the 1000-Hz region, the low-frequency slopes (Pl) of filters centered at 913 and 1095 Hz were significantly correlated at both stimulus levels, while the high-frequency slopes (Pu) were similar only at the high test level. In the 4000-Hz region, for sinusoids of 3651 and 4382 Hz, the level effect was clearer as both Pu and Pl values diverged at the low level but were related at high levels. The ERBs centered at the same CFs displayed a similar level dependence. At the stimulus level most likely to be affected by an active feedback mechanism, auditory filters centered at nearly the same frequency displayed quite distinct frequency selectivity, and this trend was stronger in the 4000-Hz region than the 1000-Hz region. The findings suggest that a saturating, active cochlear mechanism may not be distributed evenly, or contribute to peripheral tuning with equal effectiveness throughout the length of the partition. © 1997 Acoustical Society of America. [S0001-4966(97)03805-8]

PACS numbers: 43.66.Dc [WJ]

INTRODUCTION

A primary function of the mammalian ear is to analyze the spectral components of complex sounds produced by vibrations propagated through an appropriate medium. As Gold and Pumphrey (1948) pointed out, a nervous system comprised of cells that could instantaneously encode acoustic occurrences would obviate the need for a sensory organ that analyzed frequency. Because the nervous system is able to transmit a finite amount of information per unit time, and the pressure fluctuations of an acoustic signal are often too rapid to be faithfully encoded by the nervous system, there is a distinct need for the sensory organ, in this case the cochlea, to analyze a significant portion of the acoustic energy's spectrum.

Although Gold (1948) indicated that the sharp tuning of the healthy cochlea required an active component, several points remain unclear regarding the nature of the "active-ness." First, it would be beneficial to know approximately the amount of energy provided by the feedback mechanisms relative to some input level. Second, it is possible that, because structures along the cochlear partition vary morphologically in basal and apical regions, the relative contributions of an active feedback system would differ when tuning is measured across areas of the cochlea. Third, as the active processing requires an external energy source, the structures that utilize the energy may lose their active properties when damaged or otherwise metabolically compromised.

Addressing the first point is beyond the scope of this research and is currently a source of debate among research-

ers (e.g., Allen and Fahey, 1992). The second issue, as with the first, has a direct impact on the frequency-analyzing power exhibited throughout the cochlea and may be investigated psychophysically by testing a listener's performance on a frequency-resolution task that reflects, to some degree, the integrity of the cochlea's active process. The third point suggests that the active mechanical response is physiologically vulnerable and that its loss at discrete places along the cochlear partition could produce auditory filters that differ markedly. Moreover, Mills and Rubel (1996) demonstrated the contribution such mechanisms contribute to the production of distortion-product otoacoustic emissions (DPOAEs). The research reported here focused on the latter two points.

The active feedback thought to exist along the cochlear partition is evinced by a nonlinear component superimposed upon the otherwise passive motion of the basilar membrane, and level dependencies clearly exist in the behavior of this nonlinear component. Zwicker (1979) characterized the active process as one that saturated at input levels around 40 dB, and physiological measurements of the cochlear partition's motion have shown that the nonlinearity's relative contribution to the overall basilar-membrane motion is smallest at high stimulus levels. Psychophysically, such changes may be observed as the widening of auditory filters and concurrent decreases in the slopes of the auditory filters' low-frequency skirts (e.g., Formby, 1990; Moore and Glasberg, 1983; Patterson, 1976; Rosen and Baker, 1994; Rosen and Stock, 1992; Sommers and Humes, 1993; Weber, 1977). However, the effect of stimulus level at different center frequencies (CFs) is less clear. Although Weber (1977) demonstrated that the auditory filter is affected by stimulus level to essentially the same degree across several CFs each separated by an octave, Rosen and Stock reported that, as CF was

^{a)}Portions of this paper were presented at the 130th meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **98**, 2945 (A) (1995)].

raised from 125 to 1000 Hz, an increase in masker level produced a relatively greater change in filter parameters. Rosen and Stock attributed their findings to the possibility that the cochlea's nonlinear behavior was less pronounced at the apex. It should be noted that filters centered at adjacent (i.e., $<1/3$ octave) CFs have not been compared as extensively as those centered an octave apart.

A study comparing filter selectivity at neighboring CFs could address the level dependence of the active feedback process and its relative contribution at different places along the cochlear partition. Two auditory filters may exhibit similar skirt slopes at one test level and therefore filter sound with similar effectiveness. At a different intensity the slopes may diverge from one another on the low- and/or high-frequency side. The level manipulation would change cochlear tuning by altering the relative contribution of a saturating active process whose presence affects basilar membrane damping primarily in a narrow region around the location of a traveling wave's maximum displacement (e.g., Dallos, 1992). If cochlear amplification is distributed unevenly throughout the length of the partition, a manipulation of test level may exert similar, but not identical, effects on tuning at different CFs.

It is assumed that passive basilar membrane mechanics are more resistant to physical trauma than the active feedback process theorized by Gold (1948). Evidence for the vulnerability of active mechanics has been demonstrated using physiological techniques (e.g., Rhode, 1971, 1977; Robles *et al.*, 1986), and the measurement of DPOAEs (e.g., Whitehead *et al.*, 1992). Additionally, Subramaniam *et al.* (1994) indicated that the physiological vulnerability of DPOAEs varied across frequency region when measured before, during, and after a regimen of noise exposure. Specifically, after it was abolished by intense broadband noise, DPOAE amplitude returned to pre-exposure levels in the 1000-Hz region, but not the 4000-Hz region, despite total outer hair cell loss at both places. Emissions were particularly susceptible to the effects of noise when evoked by low-level primaries at the 4000-Hz place. One conclusion drawn from this work was that the outer hair cells contributed relatively more to the production of emissions in the 4000-Hz region than in the 1000-Hz region.

Psychophysical experiments have focused on the derivation of auditory-filter parameters at one or more frequencies, but the CFs examined typically differed by at least an octave. It is suggested that filter parameters derived using low-level maskers should be more dependent upon the active mechanics than those obtained with high-level maskers. Differences in processing across *slightly* different CFs, then, may provide additional evidence of an active feedback system if proximate-CF filters differ only at the low levels.

I. METHOD

A. Subjects

Auditory-filter shapes were determined for 20 adult humans [13 females and 7 males; ages 21 to 40 years old (mean = 26.8 yr)] who reported no significant exposure to occupational or recreational noise, and no regular medication with

ototoxic drugs. All listeners exhibited absolute sensitivity within 15 dB of the ANSI (1989) standard for air-conduction thresholds when screened at octave frequencies from 250 through 8000 Hz.

An additional test using a Bekesy sweep-frequency technique was conducted to assess sensitivity at interoctave frequencies, such as those at which the auditory filters were centered. The Bekesy recording was completed by 15 of the test subjects and confirmed normal absolute sensitivity. The remaining five subjects were not available for testing when the Bekesy procedure was in operation. The Bekesy tracking was conducted at a one octave/minute rate and so a determination of threshold microstructure was not made (e.g., Long, 1984).

Subjects' middle ear status was screened prior to the first test session. The screening procedure included only a measure of tympanic membrane compliance and all test subjects demonstrated normal middle ear function (e.g., Margolis and Shanks, 1985). Acoustic reflexes were not evaluated, as the central nervous system component attributed to normal reflex function was not of interest in this research.

Test ear was determined by screening subjects for spontaneous otoacoustic emissions (SOAEs) in an effort to preclude the possibility that such an emission could interact with a very low-level signal at a test frequency (e.g., Pasanen *et al.*, 1987). Two potential subjects were excluded due to the presence of SOAEs bilaterally within 250 Hz of at least one of the CFs. Four subjects who participated in the study displayed SOAEs at or near a CF in the left ear. These individuals were then screened in the right ear, which, when it did not display SOAEs, was used as the test ear. The left ear was used in all remaining subjects.

B. Instrumentation

Masked thresholds were determined with the subjects seated in a sound-treated booth and listening through TDH-49 headphones mounted in MX/41-AR cushions. Tonal signals and maskers were generated digitally at a sampling rate of 25 kHz using a Tucker-Davis Technologies QDA1 board with QAP1 array processor. Test signals were low-pass filtered at 5 kHz, while maskers were low-pass filtered at 10 kHz using a TTE1 1629 FSB filter system. Roughly one-half of the subjects were run with the masking noise filtered using a Rockland 751A Brickwall. The latter filter malfunctioned during data acquisition, prohibiting its use with all subjects and requiring two listeners, JL and TUM, to repeat 3 and 12 runs, respectively. The TTE filter was used for both signal and masker filtering following the Rockland malfunction.

The generation and presentation of the stimuli and the tracking of subjects' responses were accomplished by a personal computer (PC). Its monitor and keyboard were located inside the test booth.

C. Stimuli

Tonal signals at frequencies of 913, 1095, 3651, and 4382 Hz were presented in a simultaneous-masking context.

TABLE I. The four different masker cutoffs (LP=low-pass cutoff; HP-high-pass cutoff, both in Hz). A broadband noise masker (0–10 kHz) was also used to establish the filters' level of operation.

| Filter CF | CF(± 0.2) LP/HP | CF(± 0.4) LP/HP | CF($-0.2, 0.4$) LP/HP | CF($-0.4, 0.2$) LP/HP |
|-----------|--------------------------|--------------------------|----------------------------|----------------------------|
| 913 Hz | 730/1096 | 548/1461 | 730/1461 | 548/1096 |
| 1095 Hz | 876/1314 | 657/1533 | 876/1533 | 657/1314 |
| 3651 Hz | 2920/4391 | 2189/5121 | 2920/5121 | 2189/4391 |
| 4382 Hz | 3486/5238 | 2609/6115 | 3486/6115 | 2609/5238 |

These frequencies were selected because, in a related study (Fagelson and Champlin, 1996), they served as primary tones during the measurement of DPOAEs. Signals and maskers were synchronously gated on and off with total durations of 400 ms, including 10-ms rise–fall times.

Masker conditions included a broadband (0–10 kHz) noise, which was used in both the growth-of-masking and the auditory-filter shape tests, and four additional conditions in which symmetric and asymmetric notches were imposed on the noise. Notch-noise conditions included the cut-off frequencies indicated in Table I.

Masking noise was used to determine the auditory-filter shape at each test frequency as masked thresholds were fitted to the roex (p, r) model developed by Patterson (1976). Stone *et al.* (1992) have demonstrated a five-point roex method in which fewer maskers than were used by Patterson produced essentially the same measures of filter bandwidth and symmetry. Therefore, in the current research, a degree of resolution in the delineation of auditory-filter shapes was sacrificed for the time required to test a large pool of listeners.

D. Procedure

Signal level was adapted using a two-interval, forced-choice procedure with feedback to target 70.7% correct detection of the signal (Levitt, 1971). Subjects were presented with a visual prompt (“ready”), two intervals marked by signal lights, and a “respond” prompt. Subjects were allowed as much time as they needed to respond. Subjects responded with a keystroke appropriate to the interval in which they heard the test signal. In all testing, the mean of at least three runs was used to estimate threshold. Each run consisted of 50 trials and was retained only when the standard deviation (s.d.) within that run was less than 4.0 dB. The standard errors across the mean of runs for each masker condition, from which auditory-filter shape parameters were calculated, were less than 3.0 dB. For those conditions producing standard errors greater than 3.0 dB, at least two additional runs were completed in order to decrease within-condition variance to the 3.0-dB criterion level. Additional runs were necessary in at least one and not more than five out of the ten conditions completed by all subjects. The broadband condition was tested first, the symmetric-notch conditions second, and the asymmetric-notch conditions last. More than 75% of all conditions resulted in standard errors of <2.0 dB.

Prior to the acquisition of auditory-filter shape information, a growth-of-masking function was obtained using sig-

TABLE II. Mean ($N = 20$) relative levels of operation for the auditory filters centered at the four primary-tone frequencies, and two primary-tone levels. The target primary-tone levels were 45 and 70 dB SPL.

| Auditory filter CF | Low level | s.d. | High level | s.d. |
|--------------------|-----------|------|------------|------|
| 913 Hz | 44.94 | 2.31 | 70.23 | 2.74 |
| 1095 Hz | 45.56 | 2.15 | 70.64 | 3.19 |
| 3651 Hz | 45.33 | 2.56 | 71.74 | 3.61 |
| 4382 Hz | 46.97 | 1.98 | 72.99 | 3.21 |

nal frequencies of 1000 or 4000 Hz presented simultaneously with broadband noise. Pressure spectrum levels for these maskers were 0, 10, 20, and 30 dB SPL/Hz. A regression line was fitted to the four masked thresholds obtained from each listener at each signal frequency. Using the line's slope and y intercept, the masker levels were adjusted for each subject to produce broadband masked thresholds of either 45 dB SPL or 70 dB SPL. These thresholds were equated across subjects and were considered outputs of auditory filters (e.g., Rosen and Baker, 1994) in the wideband masker, or reference condition. The ranges of spectrum levels were 24–32 and 45–56 dB in the 1000-Hz region and 18–26 and 41–49 dB in the 4000-Hz region for the low and high levels, respectively. Obtaining the growth-of-masking functions familiarized the subjects with the test paradigm, providing at least four hours of practice time for each subject. Additionally, about half the subjects had never participated in a psychophysical experiment, and these persons practiced for two hours prior to the growth-of-masking test. Mean broadband-masked thresholds are listed in Table II and were, in 66 out of the 80 frequency-by-level conditions (82.5%), within 3 dB of the desired filter output levels.

E. Data analysis

The slopes of auditory-filter skirts [designated in the roex (p, r) procedure as Pl and Pu for the low- and high-frequency skirts, respectively] were determined using a program written for the PC by Glasberg and Moore (1990). The Pl and Pu of the filters in the 1000-Hz region are depicted schematically in Fig. 1. In the slope correlations, a slope of one of the filter's skirts, say Pl , is measured at two neighboring CFs (e.g., 913 and 1095 Hz) and the two values are correlated across the 20 subjects. The same comparison is made for the Pu values, as well as both sets of skirt slopes in the 4000-Hz region. Equivalent rectangular bandwidth (ERB) of the auditory filter is calculated from the slope information, and comparisons are made between ERBs of filters centered at neighboring CFs at the two stimulus levels.

An additional filter parameter, symmetry of the auditory filter, was determined by dividing the high-frequency slope by the low-frequency slope, producing the so-called symmetry index ($SI = Pu/Pl$) (e.g., Patterson *et al.*, 1982; Sommers and Humes, 1993). An SI of 1.0 corresponds to a filter whose low- and high-frequency slopes are identical, consistent with a symmetric filter shape. A symmetry measure of less than 1.0 indicates the low-frequency slope of the filter is steeper than the high-frequency slope, while a quotient greater than

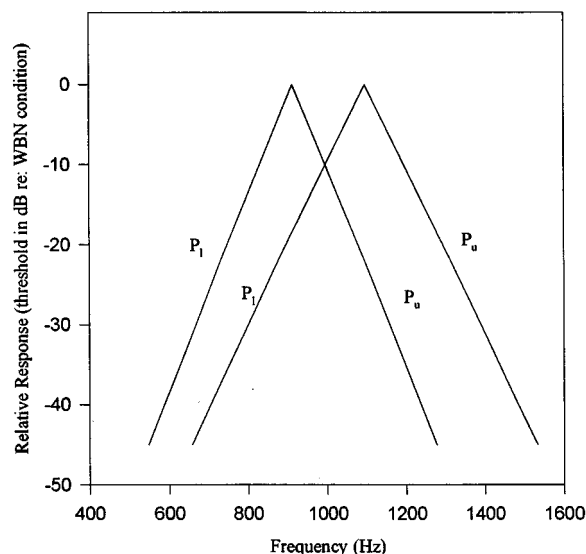


FIG. 1. Schematic representation of neighboring auditory filters centered in the 1000-Hz region. The correlations between the two P_l values are calculated in the 1000-Hz region and the 4000-Hz region, at both masker levels. The correlations between the P_u values are also determined. Additionally, the symmetry index (SI) is determined by obtaining the quotient P_u/P_l for each of the filters.

1.0 indicates a filter whose high-frequency slope is steeper than its low-frequency slope.

II. RESULTS

A. Low- and high-frequency slopes of the auditory filter

Mean slopes and standard deviations of the auditory filters centered at the four CFs appear in Table III. Additional reanalyses (see below) of slope data appear in the Appendix along with the mean values from Table III for comparison. Larger values indicate steeper slopes. Note that P_l decreased substantially, while P_u increased slightly with level. A four-way analysis of variance (ANOVA) (frequency region, CF, level, slope) using the Geisser–Greenhouse correction (GGF) for the repeated measures obtained from each subject, revealed that no main effects were significant. That is, regardless of frequency region tested, the filter slopes behaved in a similar manner when assessed at the different levels and CFs. Therefore, the slope values could be collapsed across frequency to provide the overall mean data presented in Fig. 2. The figure shows that at the low level, the low-frequency

TABLE III. Mean ($N=20$) low- and high-frequency auditory-filter slopes with one standard deviation (P_l and P_u , respectively).

| | 913 Hz | | 1095 Hz | | 3651 Hz | | 4382 Hz | |
|------------|--------|-------|---------|-------|---------|-------|---------|-------|
| | P_l | P_u | P_l | P_u | P_l | P_u | P_l | P_u |
| Low level | | | | | | | | |
| Mean | 29.2 | 22.8 | 26.5 | 26.2 | 28.3 | 24.8 | 32.4 | 25.7 |
| s.d. | 4.66 | 4.83 | 4.53 | 5.59 | 4.25 | 5.91 | 7.50 | 5.78 |
| High level | | | | | | | | |
| Mean | 24.9 | 26.7 | 21.3 | 29.1 | 21.6 | 28.1 | 21.4 | 30.5 |
| s.d. | 4.04 | 5.82 | 3.90 | 6.88 | 4.24 | 6.54 | 4.00 | 6.67 |

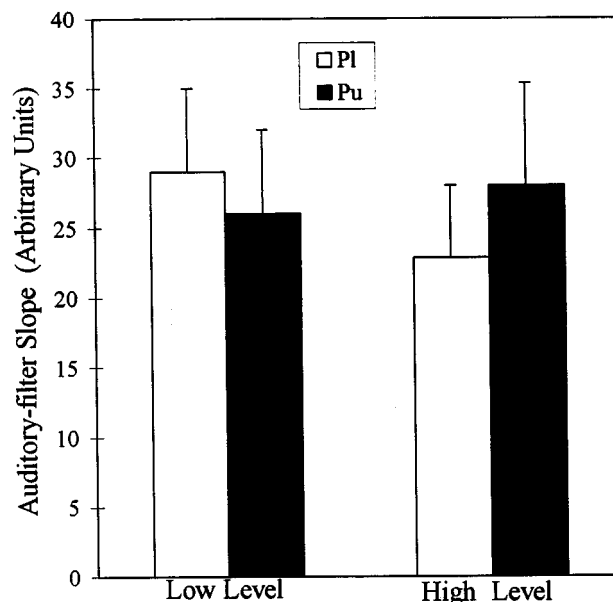


FIG. 2. Mean auditory-filter slopes assessed in both the 1000-Hz and 4000-Hz regions. The open bars reflect low-frequency slopes for all filters, the filled bars represent the high-frequency slopes for all filters. The x axis indicates the two level conditions. Error bars indicate one standard deviation.

slope is steeper than the high-frequency slope, while the reverse is true for the two slopes at the high level. The implications for filter symmetry are identified below.

The ANOVA revealed three significant interactions, one of which included the level variable. This was the interaction between level and slope of the auditory filter illustrated in Fig. 2, which was by far the strongest effect [GGF (1,19) = 2220.25; $p < 0.0001$]. The level effect is well-documented in studies of auditory-filter slope, regardless of the procedure used to derive the filter (e.g., Formby, 1990; Rosen and Baker, 1994; Weber, 1977). In the present research, the effect was associated with the slope tested, and the low-frequency slopes were affected more than the high-frequency slopes.

The remaining two significant interactions were associated with the small change to the filter centered at 913 Hz when level was changed. The two interactions, CF \times slope [GGF(1,19) = 6.262; $p < 0.05$] and frequency region \times CF \times slope [GGF(1,19) = 10.551; $p < 0.01$], reflected the finding that P_l and P_u values of the 913-Hz filter changed by nearly the same amount when masker level was increased. It is not clear why the 913-Hz CF filter behaved differently from the other filter in its region (CF = 1095 Hz).

Table IV contains the correlations between slopes of the auditory-filter skirts at neighboring CFs (i.e., P_l at 913 Hz with P_l at 1095 Hz at the low stimulus level; P_l at 3651 Hz with P_l at 4382 Hz at the low stimulus level, etc.) for each frequency and level condition. The correlations were considered better measures of filter-slope similarity than a comparison of P_l and P_u mean values at the neighboring CFs because the variability created a situation in which most mean slopes were indistinguishable. Eight correlations are reported, each depicting a measure of the similarity between

TABLE IV. Correlations tested in the two frequency regions. The association is measured between Pl values at each stimulus level at neighboring CFs. The same comparison is made for the Pu values.

| | 1000 Hz | | 4000 Hz | |
|------------|-------------------|-------------------|-------------------|-------|
| | Pl | Pu | Pl | Pu |
| Low level | 0.53 | 0.24 | -0.01 | -0.18 |
| High level | 0.64 ^a | 0.70 ^a | 0.83 ^a | 0.44 |

^a $p < 0.01$.

either set of slope values in the two frequency regions at the two levels. Statistical significance ($p < 0.01$ at $r = \pm 0.55$; degree of freedom=19) was found for three comparisons at the high masker level.

The conservative alpha level was chosen to reduce the possibility that a type I error would occur due to the number of correlations (eight) calculated with the same variables (Pl and Pu) (e.g., Glantz and Slinker, 1990). The correlations were always stronger at high signal levels than at low levels for each of the frequency regions. A significant correlation indicated that the respective slopes (Pl or Pu) for the two frequencies were closely related, or filtered the stimuli with about the same effectiveness. Thus, at the higher intensity, the auditory filters centered at neighboring frequencies assumed a more uniform shape than when assessed at the low intensity, particularly on the low-frequency side of the filter.

B. Auditory-filter ERB

The mean ERBs, expressed as a proportion of CF, appear in Table V. As for slope values, data from the reanalyses (see below) appear in the Appendix with the mean values from Table V for comparison. With the exception of CF=913 Hz, ERBs were larger at high test levels. A two-factor, repeated-measures ANOVA indicated the differences across level were significant at the remaining three CFs [GGF(1,19)=6.19; $p < 0.05$]. No main effect for frequency was found. In general, the filter bandwidths reported here were slightly wider than those reported elsewhere.

A series of correlations similar to that described above for the slope data was then conducted on the ERBs in each frequency region. Here, for example, the CF=913 Hz ERB at the low test level was compared to the CF=1095 Hz ERB at the same level for each of the 20 subjects. The same associations were measured in the 4000-Hz region and data from both the low and high test levels were examined. Table VI contains the correlations between filter ERBs at neighbor-

TABLE V. Auditory-filter ERBs expressed as a proportion of the filter's CF.

| | 913 Hz | 1095 Hz | 3651 Hz | 4382 Hz |
|------------|--------|---------|---------|---------|
| Low level | | | | |
| Mean | 0.161 | 0.157 | 0.157 | 0.143 |
| s.d. | 0.022 | 0.023 | 0.021 | 0.020 |
| High level | | | | |
| Mean | 0.160 | 0.171 | 0.172 | 0.166 |
| s.d. | 0.019 | 0.038 | 0.039 | 0.035 |

TABLE VI. Correlations between ERBs of auditory filters centered at neighboring frequencies.

| Stimulus level | 1000-Hz region | 4000-Hz region |
|----------------|-------------------|-------------------|
| Low level | 0.41 | 0.13 |
| High level | 0.86 ^a | 0.67 ^a |

^a $p < 0.01$.

ing frequencies. The correlations were far stronger at high test levels, similar to the trends reported above with regard to the filters' slopes.

C. Auditory-filter symmetry

Table VII contains means and standard deviations of the symmetry indices for each frequency and intensity condition. Generally, as the relative level of a filter increases, the filter becomes less symmetric; the low-frequency slope decreases and/or the high-frequency slope increases. The data reported here support previous findings relating the effect of masker level to increases in the SI (e.g., Formby, 1990; Sommers and Humes, 1993; Patterson *et al.*, 1982). At three of the four test frequencies (CF=1095 Hz was the exception) the low-frequency slope was steeper than the high-frequency slope at the low masker level, while the opposite relation was noted at the high masker level at all four CFs.

A two-factor (frequency and level) ANOVA with repeated measures compared SIs across the different test conditions. The analysis indicated that only the main effect of level was significant [GGF(1,19)=88.17; $p < 0.001$]. Auditory filters were more asymmetric at the high signal level (70 dB SPL) regardless of stimulus frequency. The frequency at which the filter was centered was not significant [GGF(1,19)=5.88; $p > 0.05$], and there was no interaction between frequency and level [GGF(1,19)=2.49; $p > 0.05$].

Two additional calculations of the slopes, ERBs, and correlations using a subset of the data were conducted in order to address reviewers' concerns that the five-point (Stone *et al.*, 1992) method might not provide a stable measure of auditory-filter parameters, particularly the slope values. In these analyses, two individual runs from each subject were randomly selected and then analyzed as previously described. Therefore, three analyses were completed; the first utilized the mean data (reported above), while the last two used individual runs. In all cases, the same trends originally reported were confirmed (see the Appendix).

TABLE VII. Symmetry index (SI) derived by obtaining the quotient Pu/Pl ($N=20$).

| | Auditory filter center frequency | | | |
|------------|----------------------------------|---------|---------|---------|
| | 913 Hz | 1095 Hz | 3651 Hz | 4382 Hz |
| Low level | | | | |
| Mean | 0.80 | 1.00 | 0.90 | 0.84 |
| s.d. | 0.19 | 0.22 | 0.26 | 0.27 |
| High level | | | | |
| Mean | 1.11 | 1.39 | 1.33 | 1.46 |
| s.d. | 0.31 | 0.34 | 0.31 | 0.36 |

III. DISCUSSION

An active feedback mechanism that utilizes a physiological analog to regenerative receivers has been theorized as a means by which the cochlea achieves its exquisite sensitivity and sharp tuning (Gold, 1948). Because cochlear nonlinearity saturates at moderate signal levels, the contribution to cochlear mechanics of such a process should be relatively greater when processing low-intensity rather than high-intensity sounds (e.g., Kim *et al.*, 1980; Rhode, 1971; Zwicker, 1979). In addition to being displaced by a greater amplitude relative to the intensity of the input, the partition's shape is characterized, at the lower stimulus levels, as having a sharp peak at the point of maximum displacement (e.g., Dallos, 1992; Davis, 1983).

A. Auditory-filter parameters

1. Relations between slopes of the auditory filters

Auditory filters widened with increasing signal or masker level and the ERB was dictated by a broadening of the low-frequency skirt. The correlations in Table IV showed that low- and high-frequency slopes were dissimilar when compared at neighboring CFs at lower masker levels, particularly in the 4000-Hz region (i.e., PI at CF=3651 Hz with PI at CF=4383 Hz). The effect that masker level exerted upon the correlation could reflect the integrity of a vulnerable active feedback mechanism that contributed to cochlear frequency analysis relatively more at low signal levels. With low input level, the cochlear partition may require energy in addition to that provided by the stimulus to analyze most effectively the spectral components of the stimulus. Undamping of the basilar membrane movement is likely aided by the motile activity of outer hair cells (e.g., Brownell, 1986; Dallos, 1992). But as outer hair cells are among those cochlear structures most vulnerable to the effects of noise or drugs, it is possible that listeners could have discrete and highly localized regions in which the active processes could not be fully realized. Recall that all subjects in this research had normal absolute sensitivity not only at octave frequencies but at the filter CFs, as determined with the Bekesy technique.

Correlations were affected more by the signal level in the 4000-Hz region than in the 1000-Hz region. As demonstrated by Bekesy (1960), the passive basilar membrane is displaced by a smaller amount at the cochlea's base than at its apex, given similar stimulus input levels. Therefore, the possibility exists that an active, nonlinear effect would be strongest in that part of the cochlea requiring the greatest gain from an external source. However, when such a mechanism's operation is compromised locally, auditory filters could display quite different powers of frequency analysis across small differences in CF.

In contrast, a relatively smaller contribution of the active feedback system (e.g., Whitehead *et al.*, 1992; Zwicker, 1979) would be available to auditory filters operating at a higher relative level. As more passive basilar membrane mechanics are responsible for cochlear frequency analysis at high levels, the filters would appear more homogeneous across CFs for any given listener because each listener would rely upon essentially uniform passive inner ear structures.

That is, auditory filters measured in neighboring areas should be shaped more similarly when less dependent upon vulnerable processing mechanisms.

2. Symmetry index

The symmetry measure was different from the correlations between slopes of the auditory-filter skirts discussed above in that, for measures of auditory-filter symmetry, the comparison was between the skirt slopes of any one filter. The stronger stimuli at each CF yielded a low PI , resulting in an asymmetric auditory filter characterized by an $SI > 1.0$. A similar finding was reported by Sommers and Humes (1993) when comparing the filters of hearing-impaired listeners to normal-hearing listeners who had hearing loss simulated by the introduction of external noise, and who were tested at a masker level comparable to the high level used in the present research (No=50 dB). Although specific hypotheses for such a difference were not indicated, the data were consistent with a situation in which the basilar membrane displacement pattern was quite asymmetric, with displacement amplitude greater on the basal side than on the apical side of the peak displacement. The basal spread of excitation that accompanied the higher stimulus level in the current study also suggested the mechanics available to sharpen tuning contributed relatively less to frequency selectivity at high test levels. This situation was evinced by the filter asymmetry at high test levels.

B. Relations between auditory-filter ERBs

In light of previous investigations, two important trends were noted here with regard to measurements of ERBs at the different CFs and levels. First, increasing masker spectrum level produced significantly wider ERBs. The differences were found at all frequencies, save 913 Hz, and were similar to those described by investigators such as Patterson (1976) and Weber (1977). Level effects in the current study were not as large as in Weber's, whose paradigm included several masker levels.

Glasberg and Moore (1990) suggested that changes in auditory-filter bandwidth as a function of masker level were due almost exclusively to changes in the low-frequency slope of the filter, or the high-frequency side of the displacement pattern. The shallower slopes found on the low-frequency side of a filter's CF offset the observed slight sharpening of the filter's high-frequency skirt or apical side of the excitation pattern. Recall that the low-frequency slopes of the neighboring auditory filters were more strongly correlated than the high-frequency slopes (see Table III). The ERB change, which was influenced by increased masking at points basal to the peak displacement, coincided with the greater similarity between PI 's found at high masker levels.

It is worth noting briefly that the auditory-filter skirts may be analogous to, or affected by, the pronounced absolute sensitivity variations found in threshold microstructure experiments (e.g., Long, 1984). Indeed, microstructure may have affected the data reported here by producing a narrow region (perhaps a high-dynamic range tip) that would appear sharply tuned because of an anomalous low unmasked

threshold. The effects of microstructure diminish with increasing signal level, as do the differences between frequency selectivity when such function is assessed at neighboring CFs. An experiment that tests more directly the association between the two phenomena may confirm such an interdependence.

The aforementioned level effects displayed as changes in auditory-filter slope values were echoed in the ERB data. A particularly appealing aspect of the roex (*p,r*) model relates to its treatment of *Pu* and *Pl*, which are allowed to vary independently of one another. One consequence of this independence produced the striking similarity of *Pl* values at the high test level without a concurrent homogenizing of *Pu* values. It was unclear whether the ERBs, derived from both *Pu* and *Pl*, would display the same level dependence when compared across neighboring CFs.

Consider first the difference in CFs in terms of ERB units (e.g. Moore and Glasberg, 1983). As pointed out by one reviewer, although the frequency ratio between neighboring CFs was the same in both regions (1.2), the difference expressed in ERB units was greater in the 4000-Hz region (1.6) than in the 1000-Hz region (1.37). It should follow that the auditory filters centered in the 4000-Hz region would be less similar than those centered at about 1000-Hz. In fact, Table VI shows just such a trend at both test levels. However, the level effect is stronger in the 4000-Hz region, similar to the findings reported for the slope analysis. It is interesting to consider that, despite the separation in ERB units, the *Pl* values in the 4000-Hz region are as well-correlated as measured here. As *Pl* corresponds to the basal side of the displacement pattern (e.g., Glasberg and Moore, 1990), the similarity of filter ERBs, particularly in the 4000-Hz region, is driven by the similarity of basilar membrane displacement basal to the point of maximum displacement.

When both filter skirts are combined to produce an ERB measure, then, there is a distinct smoothing effect that appears to reduce variability displayed by the *Pu* and *Pl* values considered individually (Wright, 1996). As such, the ERB may be considered less informative than an analysis addressing differences between *Pu* and *Pl*. It is clear from the data reported above that the effect of level on the association between filter bandwidths is produced, in large part, by the behavior of the filter's low-frequency skirt. As *Pl* reflects, albeit indirectly, the basal side of the basilar membrane's displacement, the similarity of filter parameters associated with high test level may be due, at least in part, to the level dependencies of the partition's displacement on the basal side of a signal's best place. Further, as level appears to exert a stronger influence on auditory-filter similarity in the 4000-Hz region, there is likely a frequency dependence to the level effect.

IV. CONCLUSIONS

(1) Auditory filters were found to be more asymmetric at high levels than at low levels. The level effect operated primarily by broadening the tuning of the filter's low-frequency slope.

(2) In both the 1000-Hz and 4000-Hz frequency regions, *Pl*, *Pu*, and ERB were more similar across filter CF

at high masker levels. This finding was attributed to the relatively smaller contribution of the active feedback mechanism apparent at high test intensities. The physiological vulnerability of such a mechanism is clear, and damage in the cochlea displaying normal absolute sensitivity may be restricted to a small area of the partition. When compared to an adjacent area, the difference in filtering ability may be quite obvious.

ACKNOWLEDGMENTS

B. R. Glasberg and B.C.J. Moore provided the software used in the fitting procedure. F. N. Martin, D. McFadden, D. C. Teas, and L. M. Thibodeau reviewed earlier drafts of this paper, and their efforts are still greatly appreciated. The text benefited substantially from the comments of Sid P. Bacon and an anonymous reviewer. These data were collected as part of a doctoral dissertation completed by the first author at the University of Texas at Austin. This work was supported (in part) by research Grant No. DC01419 from the National Institute on Deafness and Other Communication Disorders, a Jamail Grant from the University of Texas College of Communication, and by a generous contribution from the SER-TOMA Club of Northwest Austin.

APPENDIX

Mean (*N*=20) low- and high-frequency auditory-filter slopes with one standard deviation (*Pl* and *Pu*, respectively). Data from the reanalyses were derived from subsets of the mean data.

| | 913 Hz | | 1095 Hz | | 3651 Hz | | 4382 Hz | |
|--------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | <i>Pl</i> | <i>Pu</i> | <i>Pl</i> | <i>Pu</i> | <i>Pl</i> | <i>Pu</i> | <i>Pl</i> | <i>Pu</i> |
| Low level | | | | | | | | |
| Mean | 29.2 | 22.8 | 26.5 | 26.2 | 28.3 | 24.8 | 32.4 | 25.7 |
| Repeat No. 1 | 30.2 | 24.9 | 26.5 | 26.0 | 31.1 | 25.1 | 34.0 | 26.3 |
| Repeat No. 2 | 28.4 | 23.1 | 26.6 | 24.7 | 29.2 | 24.8 | 33.3 | 24.9 |
| High level | | | | | | | | |
| Mean | 24.9 | 26.7 | 21.3 | 29.1 | 21.6 | 28.1 | 21.4 | 30.5 |
| Repeat No. 1 | 25.3 | 25.8 | 21.1 | 31.0 | 21.3 | 28.0 | 21.2 | 29.4 |
| Repeat No. 2 | 24.9 | 26.8 | 21.3 | 32.3 | 22.2 | 29.3 | 20.6 | 31.1 |

Mean auditory-filter ERB reanalyses (expressed as ERB/CF).

| | 913 Hz | 1095 Hz | 3651 Hz | 4382 Hz |
|--------------|--------|---------|---------|---------|
| Low level | | | | |
| Mean | 0.161 | 0.157 | 0.157 | 0.143 |
| Repeat No. 1 | 0.156 | 0.159 | 0.156 | 0.150 |
| Repeat No. 2 | 0.163 | 0.165 | 0.160 | 0.153 |
| High level | | | | |
| Mean | 0.160 | 0.171 | 0.172 | 0.166 |
| Repeat No. 1 | 0.165 | 0.167 | 0.176 | 0.171 |
| Repeat No. 2 | 0.161 | 0.170 | 0.172 | 0.172 |

Allen, J. B., and Fahey, P. F. (1992). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," *J. Acoust. Soc. Am.* **92**, 178-188.

ANSI (1989). ANSI S3.6-1989, "American National Standard specification for audiometers" (American National Standards Institute, New York).

Bekesy, G. von (1960). *Experiments in Hearing* (McGraw-Hill, New York).

Brownell, W. E. (1986). "Outer hair cell motility and cochlear frequency selectivity," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 109-120.

- Dallos, P. (1992). "The Active Cochlea," *J. Neurosci.* **12**, 4575–4585.
- Davis, H. (1983). "An active process in cochlear mechanics," *Hear. Res.* **9**, 79–90.
- Fagelson, M. A., and Champlin, C. A. (1996). "Distortion-Product otoacoustic emissions and auditory-filter shape," Poster presentation at the 19th annual mid-winter meeting of the Association for Research in Otolaryngology, St. Petersburg, FL, 1996.
- Formby, C. (1990). "Triangular approximations of auditory-filter shape," *J. Speech Hear. Res.* **33**, 530–539.
- Glantz, S. A., and Slinker, B. K. (1990). *Primer of Applied Regression and Analysis of Variance* (McGraw-Hill, New York).
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory-filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Gold, T. (1948). "Hearing. The physical basis of the action of the cochlea," *Proc. R. Soc. Edinburgh* **135**, 492–498.
- Gold, T., and Pumphrey, R. J. (1948). "Hearing. The cochlea as a frequency analyzer," *Proc. R. Soc. Edinburgh* **135**, 462–491.
- Kim, D. O., Molnar, C. E., and Matthews, J. W. (1980). "Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve responses and in ear-canal sound pressure," *J. Acoust. Soc. Am.* **67**, 1704–1721.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 476–477.
- Long, G. R. (1984). "The microstructure of quiet and masked thresholds," *Hear. Res.* **15**, 73–87.
- Margolis, R. H., and Shanks, J. E. (1985). "Tympanometry," in *Handbook of Clinical Audiology*, edited by J. Katz (Williams and Wilkins, Baltimore), pp. 438–475.
- Mills, D. M., and Rubel, E. W. (1996). "Development of the cochlear amplifier," *J. Acoust. Soc. Am.* **100**, 428–441.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Pasanen, E. G., Wier, C. C., and McFadden, D. (1987). "Reciprocal relation between the growth of an emitted cubic distortion product and the suppression of a spontaneous otoacoustic emission," *J. Acoust. Soc. Am. Suppl.* **1** **81**, S8.
- Patterson, R. D. (1976). "Auditory-filter shapes derived with noise stimuli," *J. Acoust. Soc. Am.* **59**, 640–654.
- 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 thresholds," *J. Acoust. Soc. Am.* **72**, 1788–1803.
- Rhode, W. S. (1971). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique," *J. Acoust. Soc. Am.* **49**, 1218–1231.
- Rhode, W. S. (1977). "Some observations on two-tone interaction measured with the Mössbauer effect," in *Psychophysics and Physiology of Hearing*, edited E. F. Evans and J. P. Wilson (Academic London), pp. 27–37.
- Robles, L., Ruggero, M. A., and Rich, N. C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves and response phases," *J. Acoust. Soc. Am.* **80**, 1364–1374.
- Rosen, S., and Baker, R. J. (1994). "Characterising auditory filter nonlinearity," *Hear. Res.* **73**, 231–243.
- Rosen, S., and Stock, D. (1992). "Auditory filter bandwidths as a function of level at low frequencies (125 Hz–1000 Hz)," *J. Acoust. Soc. Am.* **92**, 773–781.
- Sommers, M. S., and Humes, L. E. (1993). "Auditory-filter shapes in normal-hearing, noise-masked normal, and elderly listeners," *J. Acoust. Soc. Am.* **93**, 2903–2914.
- Stone, M. A., Glasberg, B. R., and Moore, B. C. J. (1992). "Simplified measurement of auditory filter shapes using the notched-noise method," *Br. J. Audiol.* **26**, 329–334.
- Subramaniam, M., Henderson, D., and Spongr, V. (1994). "The relationship among distortion-product otoacoustic emissions, evoked potential thresholds, and outer hair cells following interrupted noise exposures," *Ear Hear.* **15**, 299–309.
- Weber, D. L. (1977). "Growth of masking and the auditory filter," *J. Acoust. Soc. Am.* **62**, 424–429.
- Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (1992). "Evidence for two discrete sources of 2F1-F2 distortion-product otoacoustic emission in rabbit. II: Differential physiological vulnerability," *J. Acoust. Soc. Am.* **92**, 2662–2682.
- Wright, B. A. (1996). Personal communication.
- Zwicker, E. (1979). "A model describing nonlinearities in hearing by active processes with saturation at 40 dB," *Biol. Cybern.* **35**, 243–250.

A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing

Andrew J. Oxenham

Institute for Perception Research (IPO), P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Christopher J. Plack

Laboratory of Experimental Psychology, University of Sussex, Brighton, E. Sussex BN1 9QG, England

(Received 16 July 1996; revised 5 February 1997; accepted 14 February 1997)

This paper examines the possibility of estimating basilar-membrane (BM) nonlinearity using a psychophysical technique. The level of a forward masker required to mask a brief signal was measured for conditions where the masker was either at, or one octave below, the signal frequency. The level of the forward masker at masked threshold provided an indirect measure of the BM response to the signal, as follows. Consistent with physiological studies, it was assumed that the BM responds linearly to frequencies well below the characteristic frequency (CF). Thus the ratio of the slopes of the masking functions between a masker at the signal frequency and a masker well below the signal frequency should provide an estimate of BM compression at CF. Results obtained from normally hearing listeners were in quantitative agreement with physiological estimates of BM compression. Furthermore, differences between normally hearing listeners and listeners with cochlear hearing impairment were consistent with the physiological effects of damage to the cochlea. The results support the hypothesis that BM nonlinearity governs the nonlinear growth of the upward spread of masking, and suggest that this technique provides a straightforward method for estimating BM nonlinearity in humans. © 1997 Acoustical Society of America.

[S0001-4966(97)01706-2]

PACS numbers: 43.66.Dc, 43.66.Ba, 43.66.Sr, 43.66.Mk [WJ]

INTRODUCTION

Physiological studies of cochlear mechanics have established that the response of the basilar membrane (BM) to tones at characteristic frequency (CF) is generally nonlinear and compressive (Rhode, 1971; Sellick *et al.*, 1982; Robles *et al.*, 1986; Ruggero, 1992). Damage to the cochlea, and in particular the outer hair cells (OHCs), results in a reduction in sensitivity and a loss of compression at CF (Ruggero and Rich, 1991; Ruggero *et al.*, 1993, 1995). Many of the difficulties experienced by people with sensorineural hearing loss may be explained in terms of the physiological changes associated with damage to the cochlea. For instance, the effect of abnormal growth of loudness, or ‘loudness recruitment’ (Fowler, 1936), may be due to a loss of compression in the cochlea (Yates, 1990; Glasberg and Moore, 1992). Similarly, the deterioration of performance in some measures of temporal resolution, such as gap detection in narrow-band noise (Glasberg and Moore, 1992) and the decay of forward masking (Oxenham and Moore, 1997), can also be explained in terms of loss of BM compression. Finally, a reduction in BM sensitivity to stimuli at CF results in reduced frequency selectivity, which is also a common symptom of cochlear hearing loss (e.g., Glasberg and Moore, 1986).

Given its likely influence on perception, a measure of cochlear compression in human hearing would be of considerable value. The aim of the research reported in this paper is therefore to provide a behavioral measure of BM nonlinearity in humans. The method applied here relates to the nonlinear growth of masking (GOM) observed when a masker is

well below the signal in frequency (e.g., Wegel and Lane, 1924; Egan and Hake, 1950).

One explanation for the nonlinear GOM when the masker is well below the signal in frequency is as follows. The BM responds linearly to tones with a frequency well below CF (e.g., Sellick *et al.*, 1982). A given increase in masker level will therefore be reflected by a proportional increase in BM motion at the place with a CF corresponding to the signal frequency. In contrast, the BM response to the signal at CF is compressive, and so the signal level must be increased by more than the masker level in order to produce the same change in response at the relevant place along the BM (Oxenham and Moore, 1995). Essentially the same argument has also been applied to differences in the rate-intensity functions of auditory-nerve fibers between stimuli at CF and those well below CF (Stelmachowicz *et al.*, 1987). Accordingly, a loss of compression due to cochlear hearing loss should produce a more linear GOM function. Psychoacoustic measurements using hearing-impaired listeners support this prediction (Stelmachowicz *et al.*, 1987; Murnane and Turner, 1991; Dubno and Ahlstrom, 1995; Nelson and Schroder, 1996, 1997).

While the above theory can qualitatively account for the nonlinear growth of the upward spread of masking, there is a discrepancy between the slope of the predicted GOM function, based on physiological measurements of BM nonlinearity, and that actually observed. For instance, most recent studies of BM motion have observed growth in the response to a tone at CF of around 0.2 dB/dB or less for levels above about 40 dB SPL (Sellick *et al.*, 1982; Yates *et al.*, 1990; Ruggero, 1992; Murugasu and Russell, 1995). This leads to

the prediction that the slope of a GOM function for a masker well below the signal frequency should be five or more times steeper than the GOM slope for an on-frequency masker. For simultaneous masking, a 1-dB increase in masker level should therefore produce a 5-dB increase in signal level. Instead, a survey of previous studies by Stelmachowicz *et al.* (1987) showed that, even when the masker was an octave or more below the signal frequency, the slope of the GOM function rarely exceeded 2. The slope appears to be steepest for tone-on-tone masking (van der Heijden and Kohlrausch, 1995), but even here maximum estimates range from between 2 (Schöne, 1979) and 2.5 (van der Heijden and Kohlrausch, 1995).

The discrepancy between the physiological predictions and the psychophysical results may be related to differences in the method of stimulus presentation. In the physiological studies, the response to a single tone is measured over a range of levels. In most psychophysical masking experiments, the masker and signal are presented simultaneously. The presence of the masker, by suppressing the signal, may reduce the slope of the GOM function. Ruggero *et al.* (1992) have shown not only that the BM response to a tone at CF is reduced in the presence of a low-frequency suppressor, but also that the growth of response becomes more linear. Thus, in simultaneous masking experiments, we may be measuring the more linear response to the signal in the presence of the masker, rather than the desired response of the signal alone. This would lead to an underestimate of BM compression. The possible role of suppression has also been pointed out by Nelson and Schroder (1997).

Another factor which may lead to an underestimate of compression is the possibility that listeners combine information over a number of frequency channels when detecting a signal (e.g., Zwicker, 1970). High-level signals would stimulate a greater number of frequency channels, and so enhance detection. This in turn would produce a shallower GOM slope than would otherwise be measured. The benefit of such a cue can be reduced by adding background noise to the stimuli.

In order to avoid the problems associated with suppression, we have employed a forward-masking paradigm to study the on-frequency compression of the BM. A number of previous studies have also used forward masking, including conditions where the masker was below the signal in frequency (e.g. Kidd and Feth, 1981; Nelson and Freyman, 1984; Nelson *et al.*, 1990). However, in these and other studies, the level of the signal was rarely above 40 dB SPL. It is thought that the response of the BM is more linear at lower levels, below about 30–40 dB SPL (Sellick *et al.*, 1982; Yates *et al.*, 1990; Murugasu and Russell, 1995), and so the signal levels used in previous forward-masking studies may not have been sufficiently high to measure maximum BM compression. The experiments described below employed a very brief signal, presented close to the offset of the masker, in order to measure thresholds over a wide range of signal levels. Masker, rather than signal, level was chosen as the dependent variable, as is the case in the measurement of psychophysical tuning curves.

A similar method of using the difference in response to

an on-frequency and a low-frequency stimulus to derive the amount of BM compression has been employed by Yates *et al.* (1990) in order to transform auditory-nerve rate-intensity functions into BM input-output functions. The argument that the ratio of on-frequency to off-frequency GOM slopes provides a measure of physiological response has also been used by Stelmachowicz *et al.* (1987).

I. EXPERIMENT 1. FORWARD MASKING WITH A 6-KHz SIGNAL

A. Stimuli and procedure

The level of a sinusoidal forward masker needed to mask a 6-kHz sinusoidal signal was measured for a wide range of signal levels. The masker frequency was either 3 kHz (off-frequency condition) or 6 kHz (on-frequency condition), and had a total duration of 104 ms, including 2-ms raised-cosine onset and offset ramps. The signal was also gated with 2-ms raised-cosine ramps and had no steady-state portion, giving it a total duration of 4 ms. The silent interval between the masker offset and signal onset was 2 ms (defined in terms of the zero-points in the envelope). A high signal frequency was chosen so that the signal could be made very brief while ensuring that the 3-dB bandwidth of the signal (360 Hz) fell well within the estimated equivalent rectangular bandwidth (ERB) of an auditory filter centered at 6 kHz (approximately 675 Hz; see Glasberg and Moore, 1990). This reduces the possibility of signal detection in the presence of the on-frequency (6-kHz) masker being mediated by off-frequency components of the signal, or “spectral splatter.” Also, a brief (2-ms) masker-signal interval could be employed without the response to the masker and signal temporally overlapping in the auditory periphery near the signal place, due to ringing in the auditory filters (Duifhuis, 1973). The on-frequency masker was used in order to check, and if necessary calibrate, for any nonlinear effects of forward masking *per se*. If a linear response to the 3-kHz masker at the BM place with a CF of 6 kHz is assumed, then the ratio of the slopes of the two masking functions (on-frequency and off-frequency conditions) provides an estimate of on-frequency compression.

For the normally hearing listeners, a background noise was simultaneously presented in order to restrict “off-frequency listening” (Johnson-Davies and Patterson, 1979). The background noise may also restrict “off-time listening” (Robinson and Pollack, 1973; Oxenham and Moore, 1994): It has been suggested that neural activity due to the signal may persist after the signal’s offset, and that this activity could aid detection in the presence of a forward masker. If this were the case, then a simultaneous masker which continues beyond the offset of the signal should reduce the “audibility” of this cue. For an illustration of this hypothesis, see Oxenham and Moore (1994).

For the off-frequency condition, the white, random background noise was digitally high-pass filtered (180 dB/oct slope) with a cutoff frequency of $1.117f_s$, where f_s is the signal frequency. In this case, the spectrum level of the noise in its passband was 55 dB below the level of the signal. This resulted in high-level signals being presented at least 25 dB,

TABLE I. Spectrum level of the notched noise used in experiment 1 with a 6-kHz masker and 6-kHz signal.

| Signal level (dB SPL) | Noise spectrum level (dB SPL) |
|-----------------------|-------------------------------|
| 40 | -17.0 |
| 45 | -11.5 |
| 50 | -6.0 |
| 55 | -2.5 |
| 60 | 1.0 |
| 65 | 3.5 |
| 70 | 6.0 |
| 75 | 8.5 |
| 80 | 11.0 |
| 85 | 13.5 |

and more typically 35 dB, above the simultaneous masked threshold of the signal in the background noise alone. For the on-frequency condition, a low-pass noise (cutoff frequency $0.883f_s$, 180 dB/oct filter slope) was added to the high-pass noise. The spectrum level of the resulting notched noise was set so that it was always 30 dB below the spectrum level needed to simultaneously mask the signal alone for author/listener CP. The notched-noise spectrum level used in experiment 1 at each signal level is given in Table I. The spectrum level of the notched noise was generally lower than that of the high-pass noise, as the notched noise was a more effective masker, especially at higher levels. The background noise was gated on 50 ms before the masker onset and gated off 50 ms after the signal offset, giving it a total duration of 210 ms. Due to the limited dynamic range and the probable reduced frequency selectivity of the hearing-impaired listeners, it was not thought necessary to add background noise to their stimuli. A schematic diagram of the stimulus configuration is given in Fig. 1.

All stimuli (signals, maskers, and background noise) were generated and controlled digitally on Silicon Graphics

workstations using a 32-kHz sampling rate, and analog waveforms were created using the built-in 16-bit DACs. A trial consisted of two observation intervals, marked by lights and separated by an interstimulus interval of 500 ms. The masker was presented in both intervals, and the signal was presented randomly in either the first or second interval. Thresholds were estimated using a two-alternative forced-choice paradigm with a two-up one-down adaptive procedure that estimates the masker level at the 70.7% correct point on the psychometric function (Levitt, 1971). Each reported threshold is the mean of four estimates. Listeners sat in an IAC sound-attenuating booth. The normally hearing listeners were tested at the University of Sussex, while the hearing-impaired listeners were tested at the Institute for Perception Research (IPO). Differences in setup are described below.

1. Sussex setup (normally hearing listeners)

Thresholds were measured for signal levels between 40 and 90 dB SPL in 5-dB steps. Stimuli were presented to the right ear of the listeners via a Sony MDRV6 headset, with the headset input taken directly from the output of the computer's DAC. An earplug was inserted into the left ear to prevent the detection of the signal in that ear via acoustic (or electric) crosstalk. Responses were made via a computer keyboard and feedback was provided by a graphical display on the computer monitor. In the adaptive procedure, the masker level was initially varied with a step size of 4 dB, which was reduced to 2 dB after the first four turnpoints. The threshold estimate was taken as the mean masker level at the last 12 turnpoints.

2. IPO setup (hearing-impaired listeners)

Thresholds were measured for signal levels from about 5 dB above threshold in quiet for each listener up to a maximum of about 95 dB SPL, in 5-dB steps. Stimuli were presented to the ear with the lower threshold in quiet for the brief signal at 6 kHz. For all three listeners in this study this was the left ear. Stimuli were passed from the Silicon Graphics DAC through a programmable attenuator and a headphone buffer (Tucker Davies Technologies PA4 and HB6, respectively) and were presented via a Beyer DT990 headset. No earplug was required in the opposite ear, as any crosstalk would have been well below absolute threshold in that ear. Responses were given via a response box and feedback was provided by lights on the response box. The initial step size for the adaptive procedure was 4 dB. This value was halved after every two turnpoints until a value of 1 dB was reached, and then remained constant. A run was terminated after ten reversals with a step size of 1 dB. Threshold was defined as the median masker level at the last ten turnpoints.

Both headsets (Sony MDRV6 and Beyer DT990) were calibrated with respect to their acoustic output at 1 kHz. Responses for both headsets at 2, 3, and 6 kHz (the other frequencies used in this study) were within -1 and +2.5 dB of the response at 1 kHz. Differences in the frequency response between the headphones used in the two test centers

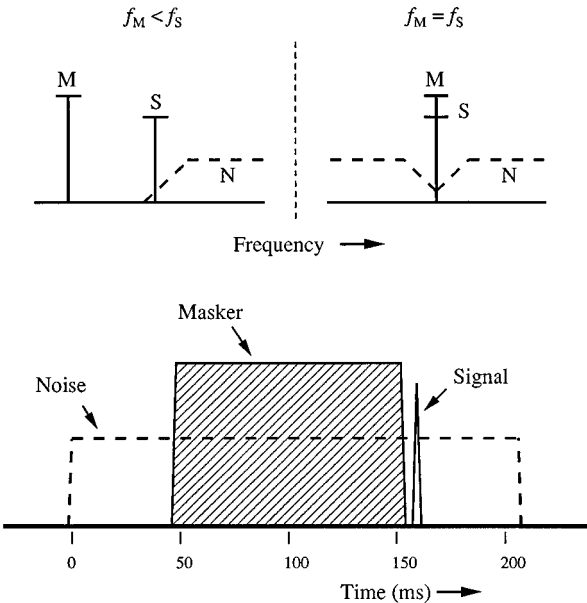


FIG. 1. Schematic diagram of the stimuli used in the experiments. The upper and lower panels show the spectral and temporal characteristics of the stimuli, respectively.

TABLE II. Pure-tone thresholds (dB HL) for the three listeners with cochlear hearing loss.

| Listener | Ear | Frequency (Hz) | | | | | |
|----------|-----|----------------|-----|------|------|------|------|
| | | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
| AR | L | 10 | 5 | 10 | 25 | 60 | 60 |
| | R | 15 | 25 | 35 | 65 | 80 | >90 |
| JK | L | 20 | 20 | 35 | 45 | 50 | 65 |
| | R | 20 | 25 | 50 | 60 | 60 | 70 |
| MV | L | 10 | 10 | 10 | 30 | 45 | 35 |
| | R | 15 | 15 | 15 | 40 | 50 | 50 |

may have affected the absolute values of masked thresholds, but should not have affected the slopes of the masking functions.

B. Listeners

Three normally hearing listeners, aged between 28 and 32, participated. One was the author CP and the others were paid an hourly wage for their services. All listeners had absolute thresholds of no more than 15 dB HL, measured at octave frequencies between 250 and 8000 Hz. Thresholds in quiet for the 4-ms signal were 28.6, 29.1, and 31.2 dB SPL for listeners CP, JB, and SD, respectively. Listeners were given at least 20 h practice before data collection began.

The three hearing-impaired listeners, MV, AR, and JK, were aged 48, 60, and 69, respectively, and were all diagnosed as having bilateral cochlear hearing loss, based on the following information: air-bone gaps were no more than 5 dB at any of the audiometric frequencies, and all three listeners had normal tympanograms and acoustic reflexes, indicating no conductive element. The normal acoustic reflexes, and the fact that speech recognition thresholds in each ear were consistent with the respective pure-tone audiometric thresholds, were interpreted as evidence against a retro-cochlear component to the hearing losses. Nothing definite

was known about the respective etiologies, but the hearing losses of AR and JK were thought to be presbycusis.

Pure-tone audiometric thresholds, measured using an Interacoustics AC-5 audiometer, are given in Table II. These thresholds were measured using long-duration tones, a modified method of limits, and a fixed step size of 5 dB. As mentioned above, the left ears of all three listeners were tested. Thresholds for the 4-ms signal, using the 2IFC adaptive procedure, were 65, 73, and 77 dB SPL for MV, AR, and JK, respectively. All three were paid for their participation and were given at least 2 h practice before data were recorded. The total practice time for the hearing-impaired listeners was therefore much less than for the normally hearing group. However, no consistent changes in the performance of the hearing-impaired listeners were noted during the course of the experiment, which lasted for a period of five weeks, with two 2-h sessions per week. This suggests that results would probably not have been different if the hearing-impaired listeners had been given more practice.

C. Results

Results are shown in Fig. 2, where masker level at threshold is plotted as a function of signal level. Note that the slope of this masking function is the reciprocal of the more usual GOM function.

Consider first the data from the normally hearing listeners (left-hand panel). Individual thresholds for the three listeners are plotted, with the error bars denoting \pm one standard deviation of the four estimates. With the masker at the same frequency as the signal (on-frequency condition; filled symbols), there are fairly large differences in threshold values for a given signal level across listeners. However, for all three listeners the masker level grows approximately linearly with signal level across the range of levels studied (cf. dashed line); the slopes of the lines fitted to the individual data are very similar, ranging from 0.96 to 1.08. Linear

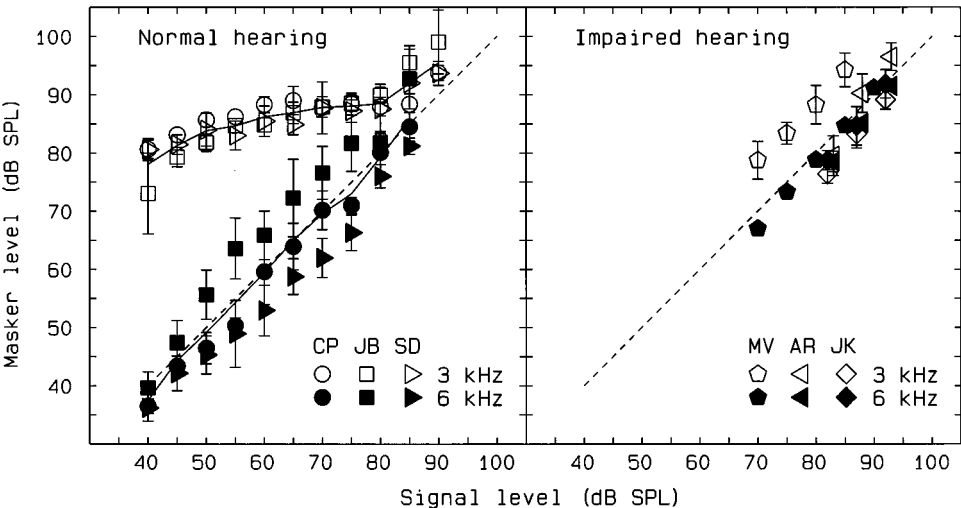


FIG. 2. The level of a masker required to mask the 6-kHz signal, as a function of signal level. Data from the three normally hearing listeners are shown in the left-hand panel and data from the three hearing-impaired listeners are shown in the right-hand panel. Error bars represent \pm one standard deviation, and are omitted if they are smaller than the respective symbol. Solid curves denote the mean thresholds of the listeners in the normally hearing group, and the dashed lines denote linear growth of masking.

growth is not usually found in forward masking (e.g., Jesteadt *et al.*, 1982); typically masker level grows more rapidly than signal level, which would lead to a steeper slope on these coordinates. However, the growth of forward masking, using a broadband forward masker and a sinusoidal signal is known to become more linear for brief signals at short masker-signal intervals (Oxenham and Moore, 1995), and the results are also consistent with a recent theory of forward masking (Oxenham and Moore, 1995, 1997).¹ In contrast to the linear masking function observed with the on-frequency masker, the results using the 3-kHz masker (open symbols) indicate highly compressive growth. For instance, increasing the signal level from 50 to 70 dB SPL results in only a 4-dB mean increase in masker level, from 83.8 to 87.8 dB SPL, across the three listeners. The slope for the mean data between signal levels of 50 and 80 dB SPL, inclusive, is 0.16. At the highest signal levels (80–90 dB SPL), growth becomes more linear for all three listeners, with individual regression lines of 0.58, 0.61, and 0.91 for listeners CP, JB, and SD, respectively. At the lowest signal levels, the masking function also seems to become more linear; individual regression lines between 40 and 50 dB, inclusive, are 0.50, 0.88, and 0.34 for listeners CP, JB, and SD, respectively.

The right-hand panel of Fig. 2 shows the individual results from the three hearing-impaired listeners.² Again, the masking function for the on-frequency masker (filled symbols) is approximately linear. This is consistent with a previous study of forward masking in hearing-impaired listeners (Oxenham and Moore, 1995). However, in contrast to the results from normally hearing listeners, masker level also grows approximately linearly with signal level in the off-frequency condition (open symbols), implying an absence of BM compression.

For two of the three hearing-impaired listeners, thresholds could only be measured for signal levels at which the masking function for the normally hearing listeners was also becoming more linear. This may mean that some of the apparent difference between the normally hearing and hearing-impaired listeners in the off-frequency condition is simply due to the exclusively high signal levels used for the hearing-impaired listeners. However, listener MV continues to show a linear masking function down to signal levels of 70 dB SPL, where the masking function for the normally hearing listeners is most compressive. Furthermore, an analysis of variance of linear regressions including only signal levels of 80 dB or more, indicated a significant difference in slope between the two groups for the off-frequency condition [$F(1,9) = 25.73$; $p < 0.001$], but not for the on-frequency condition [$F(1,7) = 0.11$; $p > 0.5$]. This implies that the difference between the normally hearing and hearing-impaired listeners in the slope of the off-frequency masking function is probably not due to the different range of signal levels used.

D. Discussion

The compression observed in the normally hearing listeners for the 3-kHz masker is, to our knowledge, greater than any previously reported in a similar task. As mentioned above, the best-fitting line for our data between signal levels

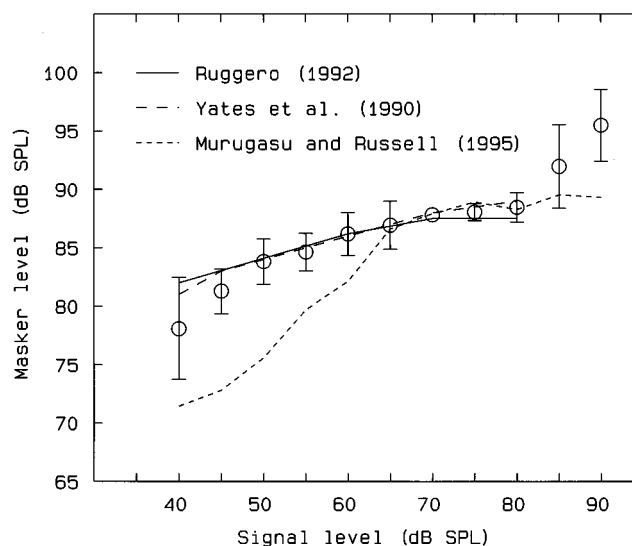


FIG. 3. A comparison of the mean 3-kHz masker data from the normally hearing listeners in Fig. 2 with physiological data of BM nonlinearity from three studies. Error bars represent \pm one standard deviation across listeners, and are omitted if smaller than the symbol. The units of dB SPL on the ordinate refer only to the psychophysical data; the absolute dB values of the physiological data are arbitrary. The characteristic frequencies in the physiological measurements were 9 kHz (Ruggero, 1992), 17.5 kHz (Yates *et al.*, 1990), and 16 kHz (Murugasu and Russell, 1995).

of 50 and 80 dB SPL, inclusive, has a slope of 0.16. In terms of GOM (signal threshold as a function of masker level), this corresponds to a slope of 6.25, which contrasts strongly with previous estimates of around 2 (e.g., Stelmachowicz *et al.*, 1987). We attribute this difference to our use of forward masking, which eliminates the effects of suppression, and the use of a very brief signal, positioned close to the masker, which allowed us to measure thresholds for signal levels higher than 50 dB SPL.

Our results may appear in conflict with those from a physiological study of Delgutte (1990). He concluded that the upward spread of masking was primarily due to the effects of suppression, while our results suggest that suppression may in fact reduce the nonlinear growth of the upward spread of masking. These apparently conflicting conclusions can, however, be reconciled if the signal levels are compared: In Delgutte's study, signal thresholds in nonsimultaneous masking conditions were rarely above 50 dB SPL, while our data show most compression above about 50 dB SPL. Thus, his conclusions may apply for low, but not high, signal levels.

Due to the linear relationship between signal level and the level of the on-frequency masker at threshold, the slope of the function for the off-frequency condition provides a direct estimate of BM compression at CF. Our estimate of 0.16 dB/dB between 50 and 80 dB SPL is in quantitative agreement with recent physiological estimates of BM compression. To illustrate this, the mean data from the normally hearing listeners in the off-frequency condition are replotted in Fig. 3, together with a sample of BM data from the chinchilla (Ruggero, 1992, Fig. 1a) and the guinea pig (Murugasu and Russell, 1995, Fig. 1b), and an estimate of BM response, derived from auditory-nerve rate-intensity func-

tions, also from the guinea pig (Yates *et al.*, 1990, Fig. 8). The BM data have been transformed from their original units (BM velocity and BM displacement) into dB and have been displaced vertically to provide easier comparison with our own data. The agreement between our data and those of Ruggero (1992) and Yates (1990) is very good. The maximum compression measured by Murugasu and Russell (1995) is similar to that of our data and the other studies, but the compression only becomes pronounced above levels of 60 dB SPL. This apparent shift of the point between more linear and more compressive response may be due to differences in the middle-ear transfer function. For instance, a horizontal shift of the function to the right, as is seen in the Murugasu and Russell data, would be the predicted result of linear attenuation due, perhaps, to a mild conductive loss in that guinea pig.

At the highest signal levels, the masking function for the off-frequency condition seems to become more linear. While some BM data show more linear growth at high levels (Ruggero and Rich, 1991), others do not (Murugasu and Russell, 1995). Unfortunately, our data cannot distinguish between these two alternatives for the following reason. If the response at CF remains compressive at high levels (beyond, say, 80 dB SPL), places along the BM with higher CFs will show a greater response to the signal than the nominal CF, in what is observed as a shift in the peak of excitation along the BM. Despite the use of background noise, we cannot rule out the possibility that, at the highest signal levels, the signal was detected at a place with a higher nominal CF and hence a more linear response to a 6-kHz tone. Another possible explanation for the apparently more linear function at high levels is that the middle-ear reflex selectively attenuates the high-level masker but not the signal, as the signal is beyond the frequency at which the reflex is effective. However, the fact that the 3-kHz masker is also probably too high in frequency for the middle-ear reflex to have any significant effect renders this explanation less likely.

At the lowest levels, the interpretation of the more linear response is also not clear-cut. Again, it is not clear from the physiological data whether, at what level, and to what extent the BM input-output function becomes more linear. In the case of our data, the steepening of the masking slope at the lowest levels does not necessarily reflect less compression. Instead, the change in slope may be due to the approach of absolute threshold. If threshold in quiet is treated as being due to a constant "internal masker," a steeper masking function is predicted as the signal approaches absolute threshold, even if the compression remains constant (Humes and Jesteadt, 1989).

In order to test whether the change in slope at low levels was due to a change in BM nonlinearity or simply the approach of absolute threshold, a further experiment using a 3-kHz masker and a 6-kHz signal was carried out with a somewhat longer signal. Increasing the duration of the signal generally increases its sensation level. Thus, at equal signal levels, the longer signal was at a higher sensation level, and so should be less influenced by the effects of absolute threshold. Second, the experiment provides a test of the generality of our hypothesis. If the shape of the masking function is

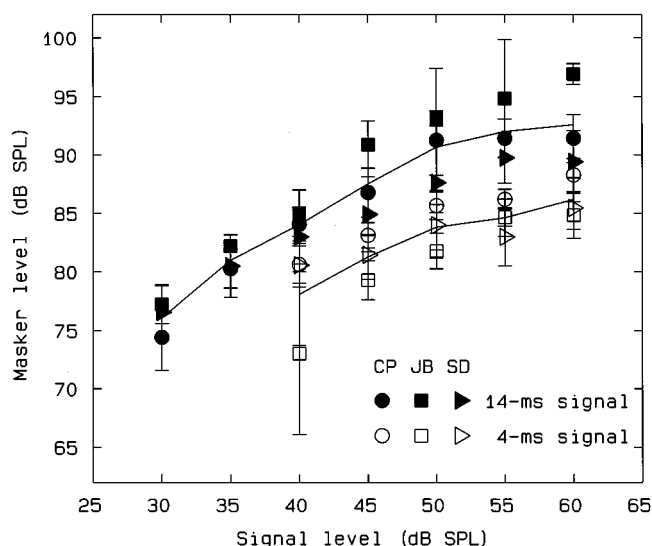


FIG. 4. Masker level at threshold as a function of signal level using a 14-ms 6-kHz signal and a 3-kHz masker for three normally hearing listeners (filled symbols). For comparison, data from Fig. 2 using a 4-ms signal are replotted as open symbols. Mean data are shown by the solid curves.

determined by BM nonlinearity, using a longer signal should result in an upwards shift of the function, but the function should be parallel to that measured with the shorter signal.

E. Experiment 1a. Effect of a longer signal

1. Method

The level of a 3-kHz masker required to mask a 6-kHz signal was measured for the three normally hearing listeners for signal levels between 30 and 60 dB SPL. The signal had a steady-state duration of 10 ms and was gated with 2-ms raised-cosine ramps. All other parameters were as described above. Thresholds in quiet for the longer signal were approximately 12 dB lower than for the original 4-ms signal.

2. Results

The results from the three normally hearing listeners are plotted as solid symbols in Fig. 4. For comparison, the data using the 4-ms signal, taken from Fig. 2, are shown as open symbols. As expected, the masker level needed to mask the longer signal is higher than for the shorter signal for a given signal level. The shape of the function matches the original function rather well. This shows first that the more linear growth below 50 dB is probably not due solely to the approach of absolute threshold and, second, that the more linear slope continues down to signal levels of 30 dB SPL.

II. EXPERIMENT 2. MEASURING COMPRESSION AT 2 KHz

The first experiment showed that normally hearing listeners exhibit strong compression at a frequency of 6 kHz. Moderate to severe (60–70 dB) cochlear hearing loss seems to eliminate the compression completely. In this experiment, we repeated part of experiment 1 at a lower frequency in order to examine whether compression varies with CF. Studies of BM nonlinearity have generally been limited to the

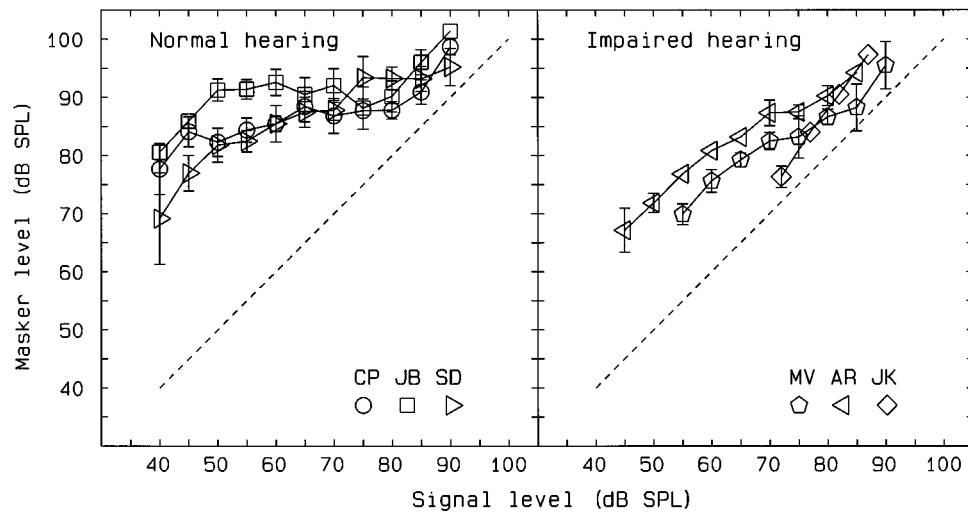


FIG. 5. Masker level at threshold as a function of signal level using a 2-kHz signal and a 1-kHz masker. Individual results from the normally hearing and hearing-impaired listeners are shown in the left- and right-hand panels, respectively. Error bars represent \pm one standard deviation, and are omitted if they are smaller than the respective symbol.

basal turn of the cochlea (highest CFs) for logistical reasons, meaning that there are essentially no reliable direct studies of BM nonlinearity at places corresponding to lower CFs. However, Cooper and Yates (1994) have derived BM input-output functions at lower frequencies from auditory-nerve rate-intensity functions in the guinea pig. They found that for fibers with CFs between 1.5 and 3.6 kHz, the derived input-output functions were more than three times steeper, and hence less compressive, than for fibers with CFs above 4 kHz.

Using the stimulus parameters as before was somewhat more problematic at 2 kHz. First, the signal bandwidth of 360 Hz is greater than that of the ERB at 2 kHz. For a masker at the signal frequency, it is difficult to rule out the detection of off-frequency components of the signal, or “spectral splatter.” This is less of a problem when the masker is below the signal in frequency, as the high-frequency slope of the masker excitation is probably shallower than that due to the splatter of the signal. Even with a masker below the signal frequency, the narrower bandwidth of the auditory filter at 2 kHz means that there is an increased chance of the masker and signal interacting on the BM, due to ringing in the filter. However, in pilot experiments, it was found that increasing the duration of the signal, or the ramps, resulted in too great a reduction in the level range over which thresholds could be measured. For this reason the temporal parameters of the stimuli were retained. The probable effect of stimulus overlap on the BM is the introduction of some suppression effects. According to our reasoning this should result in, if anything, a steepening of the slope of the masking function, and so a reduction in the apparent compression. Another possible confounding factor at 2 kHz is the influence of the middle-ear reflex. This may attenuate the 1-kHz masker more than the 2-kHz signal at high masker levels. The effect of this would again be a steeper slope and an underestimate of compression at the highest levels.

A. Method

Due to the problems of using a masker at the signal frequency, discussed above, thresholds were only measured for a 2-kHz signal and a 1-kHz masker. All other parameters were as described in experiment 1.

B. Listeners

The six listeners of experiment 1 also participated in this experiment. Thresholds in quiet for the 2-kHz, 4-ms signal were 28.3, 27.9, and 32.9 dB SPL for the normally hearing listeners CP, JB, and SD, respectively, and 44, 37, and 67 dB SPL for the hearing-impaired listeners, MV, AR, and JK.

C. Results and discussion

Figure 5 shows the results using a 2-kHz signal and a 1-kHz masker. Consider first the data from the normally hearing listeners. It can be seen that, despite the large inter-subject differences in masker level of as much as 10 dB, the slopes for all three listeners are rather flat between 50 and 80 dB SPL. We could not reliably measure masking functions for the on-frequency condition, and so it is not possible to calculate the ratio of the off-frequency slope to the on-frequency slope, as we did for the mean data at 6 kHz. For the purposes of this analysis, however, we assume that the mechanisms underlying forward masking do not change with CF and so the slope of the masking function is again interpreted as a direct measure of BM nonlinearity at 2 kHz.

The slope of the mean data from the normally hearing listeners between 50 and 80 dB is 0.17. This compares well with the slope of 0.16 for the same level range at 6 kHz. Thus, for normally hearing listeners, unlike the physiological data from the guinea pig (Cooper and Yates, 1994), our data provide no evidence for a difference in compression between CFs of 2 and 6 kHz. Again, the functions seem to become more linear at the lowest and highest signal levels.

Results from the hearing-impaired listeners are shown in the right-hand panel of Fig. 5. It was possible to measure thresholds over a much wider range of signal levels than at 6 kHz for two of the three hearing-impaired listeners (AR and MV), due to their lower absolute thresholds at 2 kHz. Again, the hearing-impaired listeners show much less compressive masking functions. For instance, between signal levels of 55 and 70 dB SPL, the slope of the mean data for listeners AR and MV is 0.75. However, the two listeners (AR and MV) with less hearing loss at 2 kHz seem to show some “residual” compression between 70 and 80 dB SPL. It seems that limited cochlear hearing loss may not necessarily result in a uniform reduction of BM compression over the entire level range, but may instead reduce the range of levels over which “normal” compression is observed. However, a larger number of listeners would be necessary to confirm this observation.

III. CONCLUDING REMARKS

The results from both experiments suggest that it is possible to derive a behavioral measure of BM nonlinearity which is in quantitative agreement with physiological measurements. We attribute the difference between these data and previous GOM data to our use of forward masking, which eliminates the effects of suppression. Another difference in stimuli between these experiments and most previous studies is the use of background noise for the normally hearing listeners. The presence of the noise may have limited the use of off-frequency and off-time listening at high signal levels, and so may have made the masking functions shallower. Cues associated with off-time listening would be the same for both on-frequency and off-frequency maskers. Also, detection of the signal’s upward spread of excitation would probably not be affected by the masker frequency. The background noise should therefore not affect the ratio between the slopes of the on- and off-frequency conditions. Thus, if our hypothesis is correct, the removal of the background noise may increase the slope of both masking functions, but the ratio, and hence the estimate of BM nonlinearity, should remain approximately constant. The background noise was not used for the hearing-impaired listeners, as in most cases the noise would have fallen below absolute threshold. However, it is possible that some of the difference observed between the two groups is due to the difference in conditions. This possibility is tested in the Appendix, using one normally hearing and one hearing-impaired listener. It is concluded there that the results cannot be accounted for by the presence or absence of the background noise.

In due course it may be possible to apply a measure similar to that used here to the diagnosis of hearing impairment. A given hearing loss may have a number of causes. For instance, selective damage to the inner hair cells (IHCs) may produce the same elevation in absolute threshold as damage to the OHCs, but may have a rather different effect on loudness perception and frequency selectivity. Patuzzi (1993) has argued that while OHC damage reduces the “active” mechanism (and hence the amplification, compression, and frequency selectivity around CF), IHC damage may produce an effect similar to a simple linear attenuation, thus

retaining compression and reasonably normal frequency selectivity. The possible perceptual consequences of such a two-component scheme are discussed in detail by Moore and Glasberg (1997), who use this approach in a model of abnormal loudness perception.

A reliable measure of BM compression may complement more traditional tests, such as loudness judgments and measures of frequency selectivity, in determining whether such a two-component approach can help account for the variability observed in hearing-impaired listeners with the same absolute hearing loss. Results from this and other studies (Oxenham and Moore, 1995; Oxenham *et al.*, 1997) indicate that compression is probably absent for hearing losses of 60 dB or more. For less severe hearing losses, however, it may be possible to test the two-component hypothesis, based on the presence and the amount of residual compression.

Some problems remain in the implementation of such tests. As discussed above, the method employed in this study is not suited to testing frequencies below about 2 kHz, which limits its use in an audiological setting. Also, the variability across the normally hearing listeners, especially in the on-frequency condition at 6 kHz (see Fig. 2), may cast doubt on the interpretation of individual data. However, the estimate of BM nonlinearity relies on the slopes of the masking functions, which were much more similar across listeners. Absolute differences in masked thresholds may not therefore play an important role. Finally the presence of the background noise had a large effect on masked thresholds, as discussed in the Appendix, although the resulting estimate of BM nonlinearity was not dependent on the presence or absence of the noise. A similarly large effect of background noise on forward-masked thresholds has been reported by Jesteadt *et al.* (1997). While the effects of noise may be accounted for by the loss of cues such as off-frequency and off-time listening, the underlying mechanisms are not quantitatively understood. An understanding of the effects of the background noise may shed light on the mechanisms involved in forward masking, and so may lead to improved techniques for measuring BM nonlinearity over a wider range of frequencies.

IV. SUMMARY

A comparison of the effects of an on-frequency forward masker with those of a forward masker well below the signal frequency provides an estimate of BM compression which is in quantitative agreement with physiological measurements. The loss of compression inferred from the results of three hearing-impaired listeners is also consistent with the physiological effects of damage to the cochlea.

The highly compressive function, derived using forward masking, suggests that suppression is not necessary for the nonlinear growth in upward spread of masking at signal levels above about 40 dB SPL. In fact, at higher signal levels, suppression may produce a more linear masking function. This may be why previous studies using simultaneous masking have found much less compression than is reported here.

If the results are accepted as providing a reasonable estimate of basilar-membrane compression, then a modified

version of the experiment may eventually be of use in the clinical diagnosis of hearing impairment.

ACKNOWLEDGMENTS

The first author is supported by a Wellcome Trust Research Fellowship (0044215/Z/95/Z), and the second author is supported by a Royal Society University Research Fellowship. We thank Armin Kohlrausch, Brian Moore, Steven van de Par, and Ian Russell for helpful comments on earlier versions of this manuscript. The constructive reviews of Marjorie Leek and David Nelson also improved the paper.

APPENDIX: EFFECTS OF BACKGROUND NOISE

For the normally hearing listeners, background noise was used to reduce the possibility of performance being improved at high signal levels by off-frequency and off-time listening. The background noise was not used for the hearing-impaired listeners because of their generally reduced dynamic range of hearing, meaning that the noise would, in most cases, have been below absolute threshold. However, to ensure that the absence of background noise was not in part responsible for the more linear masking functions found for the hearing-impaired listeners, one of the two listeners with lower thresholds at 2 kHz (listener MV) repeated some of the conditions from experiment 2 in the presence of a background noise with the same characteristics as that used for the normally hearing listeners. These results are shown in the right-hand column of Table AI, and are compared with the data from experiment 2, collected in the absence of noise. It can be seen that the presence of the noise has very little effect on performance. Thus, it is unlikely that the absence of the background noise was responsible for the more linear functions of the hearing-impaired listeners.

While the addition of the noise may reduce the slope of the masking function for normally hearing listeners, there is no *a priori* reason why the slopes from the on-frequency and off-frequency conditions should be affected differently. Therefore, while removing the noise may produce an increase in the slope of the functions, the ratio of the two slopes, and hence the estimate of BM compression, should remain roughly constant. This was tested using one normally hearing listener, CP, at the 6-kHz signal frequency. The methods and stimuli were identical to those used in experiment 1, except that the background noise was not present. The masker level necessary to mask the signal was measured for signal levels between 45 and 75 dB SPL, in steps of 10 dB.

TABLE AI. Comparison of masker levels at threshold (dB SPL) with and without background noise for hearing-impaired listener MV at masker and signal frequencies of 1 and 2 kHz, respectively. Standard deviations are shown in parentheses.

| Signal level | No noise | Noise |
|--------------|------------|------------|
| 55 | 69.9 (1.8) | 70.0 (2.0) |
| 65 | 79.3 (1.2) | 79.1 (0.3) |
| 75 | 83.2 (3.6) | 85.6 (0.9) |
| 85 | 88.3 (4.0) | 91.5 (4.6) |

TABLE AII. Comparison of results with and without background noise for listener CP. Levels are given in dB SPL. Standard deviations of the threshold estimates are shown in parentheses. An asterisk indicates that the masker level necessary was too high to be measured. The final row gives the ratio of the slopes of the off-frequency to the on-frequency regression lines.

| Conditon | Signal level | Masker level | |
|---------------------------------|--------------|--------------|-------------|
| | | No noise | Added noise |
| Off-frequency (3-kHz masker) | 45 | 84.1 (2.1) | 83.1 (1.1) |
| | 55 | 90.0 (1.2) | 86.2 (0.8) |
| | 65 | 91.8 (0.7) | 89.0 (2.4) |
| | 75 | 95.5 (0.3) | 88.4 (1.9) |
| On-frequency (6-kHz masker) | 45 | 43.4 (2.2) | 43.4 (1.7) |
| | 55 | 63.9 (7.3) | 50.3 (1.3) |
| | 65 | 87.1 (7.0) | 63.9 (4.0) |
| | 75 | * | 71.0 (1.3) |
| Ratio of slopes | | 0.17 | 0.19 |

Results are shown in Table AII, and are compared with the results from experiment 1, where noise was present. For both the on- and off-frequency conditions, the absence of noise resulted in markedly higher masker levels for a given signal level, especially at the higher signal levels. This is consistent with the idea that the background noise prevents off-frequency and off-time listening. However, a linear-regression analysis of these data showed that the ratio of the on- and off-frequency slopes remains similar both with and without noise, being slightly smaller for the no-noise condition than for the original condition, as shown in the last column of Table AII. This provides support for the idea that the ratio of the on- and off-frequency slopes can be used to estimate the amount of BM nonlinearity. The results also indicate that the strong compression observed in the experiments was probably not due to the presence of background noise.

¹Briefly, the theory postulates that the mechanisms underlying the decay of forward masking are essentially linear, but that the combination of changes in BM compression with level and the approach to absolute threshold produces a nonlinear effect overall. This scheme has been shown to be able to account for the nonlinear growth of forward masking over a wide range of masker levels, for masker-signal intervals of between 5 and 25 ms (Oxenham and Moore, 1997). The model has, however, not yet been tested on a full range of forward-masking conditions.

²For listeners MV and JK (pentagons and diamonds, respectively), the lowest signal levels measured represent 5 dB SL. For listener AR (triangles) thresholds with a 5 dB SL signal could not be measured reliably; he sometimes reported that the signal was not audible, even when the masker itself was below threshold. Repeated measures of the signal threshold in quiet indicated considerable variability. Thus, it seems likely that the 5-dB signal fell below AR's threshold in quiet at times. For this reason, only data for signal levels at 10 dB SL and above are plotted for AR.

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," *Hearing Res.* **78**, 221–234.
Delgutte, B. (1990). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," *J. Acoust. Soc. Am.* **87**, 791–809.
Dubno, J. R., and Ahlstrom, J. B. (1995). "Growth of low-pass masking of pure tones and speech for hearing-impaired and normal-hearing listeners," *J. Acoust. Soc. Am.* **98**, 3113–3124.
Duifhuis, H. (1973). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," *J. Acoust. Soc. Am.* **54**, 1471–1488.
Egan, J. P., and Hake, H. W. (1950). "On the masking pattern of a simple auditory stimulus," *J. Acoust. Soc. Am.* **22**, 622–630.

- Fowler, E. P. (1936). "A method for the early detection of otosclerosis," *Arch. Otolaryngol.* **24**, 731–741.
- Glasberg, B. R., and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am.* **79**, 1020–1033.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.* **47**, 103–138.
- Glasberg, B. R., and Moore, B. C. J. (1992). "Effects of envelope fluctuations on gap detection," *Hearing Res.* **64**, 81–92.
- Humes, L. E., and Jesteadt, W. (1989). "Models of the additivity of masking," *J. Acoust. Soc. Am.* **85**, 1285–1294.
- Jesteadt, W., Bacon, S. P., and Lehman, J. R. (1982). "Forward masking as a function of frequency, masker level, and signal delay," *J. Acoust. Soc. Am.* **71**, 950–962.
- Jesteadt, W., Wilke, S., and Bacon, S. P. (1997). "Combinations of forward and simultaneous masking," *J. Acoust. Soc. Am.* (submitted).
- Johnson-Davies, D., and Patterson, R. D. (1979). "Psychophysical tuning curves: Restricting the listening band to the signal region," *J. Acoust. Soc. Am.* **65**, 765–770.
- Kidd, G., and Feth, L. L. (1981). "Patterns of residual masking," *Hearing Res.* **5**, 49–67.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Moore, B. C. J., and Glasberg, B. R. (1997). "A model of loudness perception applied to cochlear hearing loss," *Aud. Neurosci.* **3**, 289–311.
- Murmane, O., and Turner, C. W. (1991). "Growth of masking in sensorineural hearing loss," *Audiology* **30**, 275–285.
- 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., Chargo, S. J., Kopun, J. G., and Freyman, R. L. (1990). "Effects of stimulus level on forward-masked psychophysical tuning curves in quiet and in noise," *J. Acoust. Soc. Am.* **88**, 2143–2151.
- Nelson, D. A., and Freyman, R. L. (1984). "Broadened forward-masked tuning curves from intense masking tones: delay-time and probe level manipulations," *J. Acoust. Soc. Am.* **75**, 1570–1577.
- Nelson, D. A., and Schroder, A. C. (1996). "Release from upward spread of masking in regions of high-frequency hearing loss," *J. Acoust. Soc. Am.* **100**, 2266–2277.
- Nelson, D. A., and Schroder, A. C. (1997). "Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss," *J. Acoust. Soc. Am.* **101**, 2186–2201.
- Oxenham, A. J., and Moore, B. C. J. (1994). "Modeling the additivity of nonsimultaneous masking," *Hearing Res.* **80**, 105–118.
- Oxenham, A. J., and Moore, B. C. J. (1995). "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. (1997). "Modeling the effects of peripheral nonlinearity in normal and impaired hearing," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ).
- Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (1997). "Short-term temporal integration: Evidence for the influence of peripheral compression," *J. Acoust. Soc. Am.* **101**, □□–□□.
- Patuzzi, R. B. (1993). "Otoacoustic emissions and the categorization of cochlear and retrocochlear hearing losses," *Br. J. Audiol.* **27**, 91–95.
- 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.
- Robinson, C. E., and Pollack, I. (1973). "Interaction between forward and backward masking: A measure of the integrating period of the auditory system," *J. Acoust. Soc. Am.* **53**, 1313–1316.
- Robles, L., Ruggero, M. A., and Rich, N. C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and phase responses," *J. Acoust. Soc. Am.* **80**, 1364–1374.
- Ruggero, M. A. (1992). "Responses to sound of the basilar membrane of the mammalian cochlea," *Curr. Op. Neurobiol.* **2**, 449–456.
- 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., and Recio, A. (1993). "Alteration of basilar membrane responses to sound by acoustic overstimulation," in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. v. Dijk, and S. M. v. Netten (World Scientific, Singapore), pp. 258–265.
- Ruggero, M. A., Rich, N. C., Robles, L., and Recio, A. (1995). "The effects of acoustic overstimulation, other cochlear injury and death on basilar membrane responses to sound," in *Effects of Noise on Hearing: Vth International Symposium*, edited by R. J. Salvi, A. Axelsson, D. Henderson, and R. Hamernik (Thieme Medical, Stockholm).
- Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (1992). "Basilar membrane responses to two-tone and broadband stimuli," *Philos. Trans. R. Soc. London, Ser. B* **336**, 307–315.
- Schöne, P. (1979). "Mithörschwellen-Tonmuster maskierender Sinustöne," *Acustica* **43**, 197–204.
- Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.* **72**, 131–141.
- Stelmachowicz, P. G., Lewis, D. E., Larson, L. L., and Jesteadt, W. (1987). "Growth of masking as a measure of response growth in hearing-impaired listeners," *J. Acoust. Soc. Am.* **81**, 1881–1887.
- van der Heijden, M., and Kohlrausch, A. (1995). "The role of envelope fluctuations in spectral masking," *J. Acoust. Soc. Am.* **97**, 1800–1807.
- Wegel, R. L., and Lane, C. E. (1924). "The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear," *Phys. Rev.* **23**, 266–285.
- Yates, G. K. (1990). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions," *Hearing Res.* **50**, 145–162.
- Yates, G. K., Winter, I. M., and Robertson, D. (1990). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," *Hearing Res.* **45**, 203–220.
- Zwicker, E. (1970). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 376–394.

Short-term temporal integration: Evidence for the influence of peripheral compression

Andrew J. Oxenham^{a)}

Institute for Perception Research (IPO), P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Brian C. J. Moore and Deborah A. Vickers

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 9 September 1996; accepted for publication 30 December 1996)

Thresholds for a 6.5-kHz sinusoidal signal, temporally centered in a 400-ms broadband-noise masker, were measured as a function of signal duration for normally hearing listeners and listeners with cochlear hearing loss over a range of masker levels. For the normally hearing listeners, the slope of the function relating signal threshold to signal duration (integration function) was steeper at medium masker levels than at low or high levels by a factor of nearly 2, for signal durations between 2 and 10 ms, while no significant effect of level was found for signal durations of 20 ms and more. No effect of stimulus level was found for the hearing-impaired listeners at any signal duration. For signal durations greater than 10 ms, consistent with many previous studies, the slope of the integration function was shallower for the hearing-impaired listeners than for the normally hearing listeners. However, for shorter durations, there was no significant difference in slope between the results from the hearing-impaired listeners and those from the normally hearing listeners in the high- and low-level masker conditions. A model incorporating a compressive nonlinearity, representing the effect of basilar-membrane (BM) compression, and a short-term temporal integrator, postulated to be a more central process, can account well for changes in the short-term integration function with level, if it is assumed that the compression is greater at medium levels than at low or high levels by a factor of about 4. This is in reasonable agreement with physiological measurements of BM compression, and with previous psychophysical estimates. © 1997 Acoustical Society of America. [S0001-4966(97)05606-3]

PACS numbers: 43.66.Dc, 43.66.Ba, 43.66.Sr, 43.66.Mk [JWH]

INTRODUCTION

This study examines temporal integration, or how the threshold for detecting a signal depends on signal duration, and investigates the extent to which the results may be influenced by peripheral compression in the auditory system. Although temporal integration has been the subject of intense study since the 1940s, there is still no consensus as to the underlying mechanisms involved. The fact that thresholds for sinusoidal signals in quiet or in background noise decrease by approximately 3 dB per doubling in duration, between about 10 and 200 ms, led to the hypothesis that stimulus intensity (above a certain minimum intensity) is fully integrated (Hughes, 1946; Garner and Miller, 1947). A different formulation, in terms analogous to a simple electrical RC circuit, was proposed independently by Feldtkeller and Oetinger (1956) and Plomp and Bouman (1959). Both types of model lead to very similar predictions (Plomp and Bouman, 1959), namely a 3-dB decrease in threshold per doubling of duration up to a certain duration (between about 100 and 300 ms) followed by asymptotic behavior. More sophisticated models of temporal summation, taking neural activity into account, have been proposed by Zwislöck (1960, 1969).

As previously pointed out, most recently by Viemeister and Wakefield (1991), the fact that the auditory system can

act like an energy detector over a limited duration does not necessarily imply true integration of an intensity-like quantity. Results can often be equally well described in terms of an increase in signal duration leading to an increase in the statistical probability of detection. In a recent exposition of such a theory, known as the “multiple-looks” hypothesis, Viemeister and Wakefield (1991) presented results, involving the detection of two separate tone bursts, which cannot be accounted for by a long-term temporal integrator. Another possible approach has recently been described by Dau *et al.* (1996a,b). In their model an analysis window (template) is used, which is matched to the time pattern of the signal. With an extension of the original model, Dau and colleagues have successfully modeled the data of Viemeister and Wakefield (1991) (Dau *et al.*, 1997). Whether or not long-term temporal integration is due to a long time constant, multiple looks, or an adjustable template, it is clear that there is no “hard-wired” long time constant in the auditory system that cannot be bypassed. If this were the case, then thresholds for a very brief signal, temporally centered in a masker, would increase with masker duration for durations beyond 100 ms. In a series of experiments, Penner and her colleagues showed that in fact the threshold of a brief signal pulse, temporally centered in a broadband-noise masker, increased for masker durations up to between 10 and 20 ms, and then remained roughly constant or decreased (Penner *et al.*, 1972; Penner and Cudahy, 1973). The masker duration at which thresholds

^{a)}Electronic mail: oxenham@ipo.tue.nl

ceased to increase was termed the critical masking interval. These results were interpreted as reflecting a short-term integrator with an effective time constant of around 10 ms. This is broadly consistent with many other measures of temporal resolution, such as gap and decrement detection and the decay of forward masking (e.g., Buus and Florentine, 1985; Oxenham and Moore, 1994; Peters *et al.*, 1995).

We therefore adopt the position that in cases where detection is achieved by an overall change in level, *there is a short-term temporal integrator which cannot be bypassed, and which determines thresholds for signal durations up to about 10 ms.* In this paper, we refer to this as *short-term temporal integration*. For signal durations greater than this, thresholds may be determined by a multiple-looks strategy, longer time constants, an adjustable template, or a combination of these.

Even if we accept the proposition of true temporal integration for short signal durations, it is still not necessary to assume that a quantity proportional to signal intensity is integrated. Penner (1978) has shown, for instance, that using a power-law nonlinearity prior to integration, complementary pairs of nonlinearities and temporal-weighting functions can be constructed which all produce the same time-intensity trade function. Thus the slope of the temporal-integration function is not sufficient in itself to determine what quantity is integrated. From this, Penner (1978) was able to show that, for a given temporal window, *a change in the nonlinearity leads to a change in the slope of the integration function.* Specifically, the more compressive the nonlinearity, the steeper (more negative) the slope of the function [relating signal level (dB) to log (duration)]. An intuitive explanation of this relationship is given in the Appendix.

In a number of previous studies, the compression used in this type of model has been linked to peripheral auditory compression (e.g., Oxenham and Moore, 1995; Moore *et al.*, 1996; Oxenham and Plack, 1997). This approach has been stimulated by physiological measurements of basilar-membrane (BM) motion (Rhode, 1971; Sellick *et al.*, 1982; Ruggero, 1992; Ruggero *et al.*, 1995). Essentially, the response of the BM to sound at the characteristic frequency (CF) of the place of measurement appears to be highly compressive, especially for levels between about 40 and 80 dB SPL. Damage to the cochlea reduces or eliminates compression.

In terms of a short-term integration model, the nonlinearity, representing peripheral (BM) compression, is followed by a linear integrator, representing a somewhat higher processing stage. Relating the model's compression to BM nonlinearity, and always assuming an invariant temporal-window shape, produces the following two predictions. First, for normally hearing listeners, greater mid-level BM compression should produce a steeper slope in the integration function at medium levels than at low or high levels. Second, to the extent that the BM response is more linear in listeners with cochlear hearing loss, the slope of the integration function should be shallower for hearing-impaired listeners than for normally hearing listeners.

This second prediction is extremely well documented for unmasked (absolute) thresholds (e.g., Miskolczy-Fodor,

1953; Elliott, 1963; Wright, 1968; Pedersen and Elberling, 1973; Chung, 1981; Florentine *et al.*, 1988): There is indeed in general a reduction in the slope of the temporal-integration function for listeners with sensorineural hearing loss. This is not due to the generally higher levels at which the stimuli are presented to the hearing-impaired listeners (Gengel, 1972; Florentine *et al.*, 1988) and cannot be accounted for by the possible detection of spectral splatter (Florentine *et al.*, 1988; Carlyon *et al.*, 1990). It has been previously suggested that this reduction in slope over the whole range of signal durations is due to reduced peripheral compression (Moore, 1991, 1995), but to our knowledge no quantitative test of this hypothesis has been attempted.

The first prediction of the model that, for normally hearing listeners, short-term temporal integration should have a steeper slope at medium signal levels than at higher or lower levels, has much less experimental support. A test of the hypothesis would need to fulfill the following conditions. First, the number of points measured for durations of 10 ms or less must be sufficient to give a reasonable estimate of the slope of the integration function. Second, in order to avoid problems of detection by combining information across different frequency channels, the bandwidth of the signal, even at the shortest durations, must fall approximately within one critical band. This essentially limits the choice of signal frequencies to those above about 4 kHz. Third, the masker level must be chosen so that the signal level at durations of 10 ms or less lies within the level region thought to be most compressive, namely between about 50 and 70 dB SPL (Oxenham and Plack, 1997). For comparison with regions of more linear processing, masker levels must lie well below or above that level. These restrictions severely limit comparisons within the available literature. Both Florentine *et al.* (1988) and Gengel (1972) have compared the slope of integration for a sinusoidal signal in the presence of a masking noise with that in quiet. Interestingly, they were looking for evidence that the slope decreased in the presence of a masking noise, while we expect the reverse. Unfortunately, Gengel (1972) did not measure thresholds for durations less than 10 ms, even at 4 kHz. Florentine *et al.* (1988) have a number of data points which could come into consideration in terms of signal frequency and duration. However, only one of the noise levels they used (simulation of hearing-impaired listener RT) produced thresholds below 70 dB SPL for signals of 10 ms or less at 4 kHz. In Fig. 5 of their paper, for durations less than 10 ms, there is a tendency for the slope of the masked thresholds to be steeper than that of the thresholds in quiet. However, the difference is small, is based on only three durations, and is therefore far from conclusive.

The most positive evidence for a change in the integration function with level comes from a study by Stephens (1973). In that study, performance was measured in terms of percent correct for sinusoidal signals at a constant signal-to-noise ratio (in terms of overall energy), for a number of signal durations and noise-masker levels. It was found that performance at the shortest duration (2 to 3 cycles of the signal) was strongly dependent on masker levels, reaching a minimum at medium masker levels. For signal durations of 20 ms and longer, performance was independent of masker

level. This was true for signal frequencies of both 1 and 4 kHz. The results of Stephens (1973) imply that the slope of the integration function may indeed be steeper at medium levels. However, as he measured performance for a fixed signal level, it is not possible to derive the slope of the integration function from his data. Second, as his signals were switched on and off without ramps, it is not clear to what extent the integration of signal energy across frequency played a role in his experiments.

Finally, recent experiments on the loudness of sinusoids as a function of duration have shown that the difference in loudness between a short and a long tone is greatest at medium levels (Florentine *et al.*, 1996). The authors note that the results are consistent with greater BM compression at medium levels. However, quantitative analysis of on-frequency compression based on loudness judgments is difficult, due to the fact that loudness is almost certainly influenced by off-frequency excitation (e.g., Zwicker, 1960).

In summary, while there are indications that the slope of the short-term temporal-integration function may be steeper at medium levels, the available data are neither conclusive nor sufficient for quantitative analysis. In the experiment described below, thresholds for a 6.5-kHz sinusoidal signal in a broadband-noise masker were measured as a function of signal duration in both normally hearing and hearing-impaired listeners over a range of masker levels. Based on the assumption that the shape of the temporal window remains invariant with level, and that it is not altered by cochlear pathology, the basic prediction for normally hearing listeners is that the slope of the integration function at medium levels will be steeper than at low and high levels. Assuming a linear BM input-output function, the slope of the integration function should remain constant with level for the hearing-impaired listeners. The difference between this slope and that of the normally hearing listeners at low and high levels should provide an indication of the amount of compression present in normal hearing at the lowest and highest levels, always assuming the same or similar temporal windows across the two groups.

I. EXPERIMENT 1. TEMPORAL INTEGRATION AT 6.5 kHz

A. Stimuli

The masker was a bandpass-filtered Gaussian noise (Hewlett-Packard 3722A) with cutoff frequencies of 2 and 12 kHz (Kemo VBF/8/03 filter, 48-dB/oct slope). The signal was a 6.5-kHz sinusoid (Farnell DSG1). Both masker and signal were gated with 1-ms raised-cosine ramps, and the signal was always temporally centered within the 400-ms masker. Thresholds were measured for signals with half-amplitude durations ranging from 2 to 200 ms (1–199 ms steady state). For the normally hearing listeners, thresholds were measured at masker spectrum levels of -10 , 20 , and 50 dB (*re*: $20 \mu\text{Pa}$).¹ For the hearing-impaired listeners, masker spectrum levels of 30 , 40 , and 50 dB were tested. These levels were chosen to span the dynamic range of the listeners, such that thresholds for the longest duration signal (200 ms) and the lowest masker spectrum level (-10 dB for the

normally hearing and 30 dB for the hearing-impaired listeners) were about 5 dB above individual thresholds in quiet for all listeners.

For the normally hearing listeners, stimulus timing was controlled by a Texas Instruments 990/4 computer system, and the signal level was varied using a Charybdis model D programmable attenuator. Two pairs of analog multipliers (AD 534L) in series were used as gates for the masker and signal, giving an on-off ratio exceeding 100 dB. For the hearing-impaired listeners, who were tested at a later time, stimuli were controlled using a Tucker-Davies Technologies (TDT) system with a PC. The masker and signal were gated and attenuated, using two switches (TDT SW2) and two programmable attenuators (TDT PA4), before being added (TDT SM3) and passed through a headphone buffer (TDT HB6). For both groups, the stimuli were then passed through a final manual attenuator (Hatfield 2125) to one earphone of a Sennheiser HD414 headset. For the normally hearing listeners, the stimuli were presented to the left ear. For the hearing-impaired listeners, the stimuli were presented to the ear with the lower absolute threshold at 6.5 kHz. For listeners VT and DT, this was the right ear and for listeners AW and MG this was the left ear.

A trial consisted of two observation intervals, marked by lights, separated by a silent interval of 500 ms. The 400 -ms masker burst occurred in both intervals and the signal was presented randomly in either the first or the second interval.

B. Procedure

Thresholds were determined using a two-alternative forced-choice method with a three-down one-up adaptive procedure that estimates the 79.4% correct point on the psychometric function (Levitt, 1971). The initial step size was 5 dB, which was reduced to 2 dB after the first four reversals. A run was terminated after a total of 12 reversals and the threshold was defined as the mean of the levels at the last 8 reversals. Each data point reported here is the mean of three such threshold estimates. Listeners were tested individually in a double-walled sound-attenuating chamber.

C. Subjects

Four normally hearing listeners and four listeners with cochlear hearing loss participated as subjects. Two of the normally hearing listeners were authors AO and DV, one (MS) was a member of the laboratory who volunteered for the experiment, and the other (ST) was paid an hourly wage for her participation. Audiometric thresholds for all four listeners were 15 dB HL or less for octave frequencies between 250 and 8000 Hz. The ages of the normally hearing listeners ranged from 25 to 34 years. All normally hearing listeners had extensive experience in psychoacoustic tasks and were given at least 1 -h practice before data were collected.

The four hearing-impaired listeners were selected on the basis of having a sensorineural hearing loss of between 40 and 60 dB at the test frequency (6.5 kHz). All had air-bone gaps of less than 10 dB and showed normal tympanometry, indicating no conductive element. There was no sign of tone decay for any of the four listeners (tone decay is often a

TABLE I. Summary characteristics for the hearing-impaired listeners, showing their ages, genders, diagnoses, and the audiometric thresholds for the test ears, given in dB HL.

| Listener | Age | Sex | Frequency (Hz) | | | | | | | Diagnosis |
|----------|-----|-----|----------------|-----|------|------|------|------|------|---------------|
| | | | 250 | 500 | 1000 | 2000 | 4000 | 6000 | 8000 | |
| AW | 81 | M | 5 | 0 | 5 | 15 | 45 | 50 | 60 | presbycusis |
| VT | 61 | M | 20 | 10 | 5 | 10 | 50 | 40 | 50 | noise induced |
| MG | 71 | F | 5 | 10 | 20 | 50 | 45 | 55 | 50 | presbycusis |
| DT | 72 | M | 20 | 10 | 10 | 45 | 65 | 60 | 60 | noise induced |

symptom of retrocochlear loss) and all showed recruitment, as indicated by a smaller-than-normal range between threshold and the highest comfortable level, which is a characteristic of cochlear hearing loss. Speech discrimination was not measured.²

One listener (AW) had extensive previous experience in psychoacoustic tasks. The other three listeners were given at least 4-h practice before data were collected. Audiometric thresholds for the ears that were tested, together with each listener's diagnosis, gender, and age, are given in Table I.

D. Results

Data from the individual listeners are plotted in Fig. 1. The left and right columns show data from the normally

hearing and hearing-impaired listeners, respectively. Each panel represents a different masker spectrum level, as shown in the insets. The solid curves denote mean thresholds across listeners for each condition. The data for normally hearing listener MS (diamond symbols) in the 50-dB condition have been shifted upwards by 5 dB, for ease of comparison, and are treated in the analyses below as if they had been measured in the presence of a 50-dB masker (see footnote 1).

There are some individual differences in the data, in terms of both overall sensitivity and the slope of the integration function. For instance, hearing-impaired listener MG (right panels; circles) is generally less sensitive than the other three listeners, especially at the 40- and 50-dB masker spectrum levels, and normally hearing listener ST (left panels; squares) exhibits a shallower slope of integration than the other listeners, especially in the -10 -dB condition. The shallower slope of ST does not seem to be due to an elevated absolute threshold in quiet: while ST's threshold in quiet for the 200-ms signal (14 dB SPL) was higher than that of AO (6.8 dB SPL), it was lower than that of MS (16.2 dB SPL).

Initially, single-line linear regression analyses, in terms of signal level (dB SPL) as a function of $10 \log [\text{duration (ms)}]$, were performed across all signal durations. Resulting slopes for the individual and group mean data in the different conditions are given in Table II. For the individual slopes, the three estimates for each data point shown in Fig. 1 were used. For the group mean slopes, the individual mean thresholds shown in Fig. 1 were used and converted to deviations from the mean for that condition and listener, thus compensating for differences in overall sensitivity across listeners.

In Table II it can be seen that for all the normally hearing listeners the slope of the function for the 20-dB condition is steeper than for the -10 - and 50-dB conditions. The effect of masker level is not so pronounced for the hearing-impaired listeners. A within-groups comparison of the slopes for the mean data confirmed this impression: There was no significant effect of masker level for the hearing-impaired group [$F(2,93) = 1.08$, $p > 0.3$], while for the normally hearing group, the effect of masker level was highly significant [$F(2,93) = 21.46$, $p \leq 0.0001$], reflecting the steeper overall slope of the 20-dB condition. Although the slopes from the hearing-impaired group are generally shallower than those from the normally hearing group, the difference is not as great as that sometimes reported in the literature, where a difference of a factor of 2 is not uncommon for listeners with a hearing loss of 40 dB or more (e.g., Pedersen and Elberling, 1973).

Visual inspection of the data in Fig. 1 raises some doubt

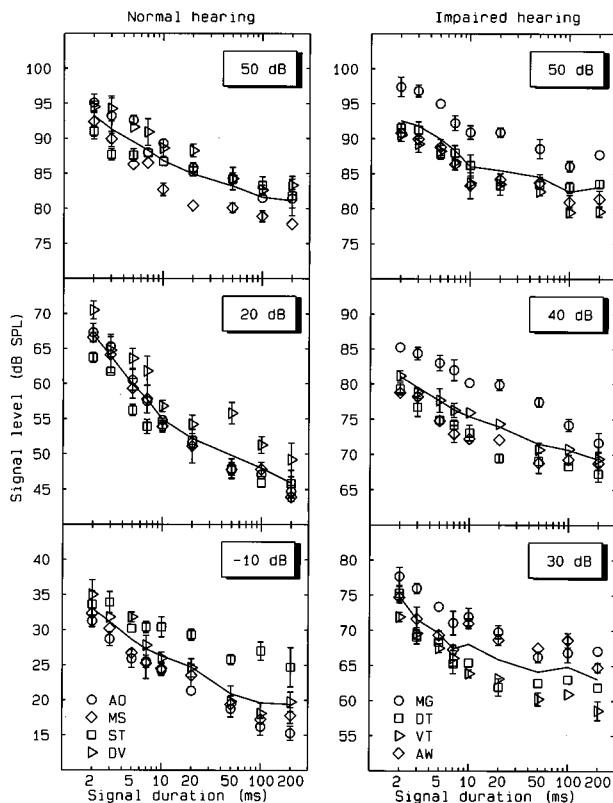


FIG. 1. Individual data from experiment 1. Signal level at threshold is plotted against signal duration on a log scale. The left and right columns represent data from the normally hearing and hearing-impaired listeners, respectively. The masker spectrum level in each condition is given in the insets. Error bars represent ± 1 standard deviation, and are omitted if smaller than the height of the symbol. The solid curves show the mean thresholds of the listeners within each condition.

TABLE II. Slopes from a single-line linear regression analysis [signal level (dB) against 10 log (duration)] of the data from Fig. 1. Masker spectrum level is in dB SPL. All slopes are negative.

| Group | Masker level | Listener | | | | Mean data |
|----------|--------------|----------|------|------|------|-------------|
| | | AO | ST | DV | MS | |
| Normal | -10 dB | 0.79 | 0.44 | 0.84 | 0.74 | 0.70 |
| | 20 dB | 1.14 | 0.89 | 0.95 | 1.09 | 0.92 |
| | 50 dB | 0.71 | 0.39 | 0.64 | 0.70 | 0.61 |
| Impaired | | MG | DT | VT | AW | |
| | 30 dB | 0.55 | 0.53 | 0.62 | 0.33 | 0.51 |
| | 40 dB | 0.66 | 0.58 | 0.57 | 0.51 | 0.58 |
| | 50 dB | 0.55 | 0.44 | 0.56 | 0.48 | 0.51 |
| | | | | | | |

as to the appropriateness of a single-line regression. Consistent with some previous data (Green *et al.*, 1957; Stephens, 1973; Florentine *et al.*, 1988), the integration function seems to be steeper at shorter durations than at longer ones. This is especially apparent in the 20-dB condition for the normally hearing group, but is also visible, for instance, in the 30- and 50-dB conditions for the hearing-impaired group. The steepening can probably not be accounted for by the spread of signal energy outside the auditory filter centered on the signal frequency, as the 3-dB bandwidth of the shortest signal (ca. 420 Hz) is less than the estimated equivalent rectangular bandwidth of the auditory filter at 6.5 kHz (ca. 725 Hz; see Glasberg and Moore, 1990). The impression of two separate regions was tested by comparing the one-line fit with that using two lines. In the latter procedure, the data in each condition were divided by duration into two subconditions. A linear regression was carried out independently on each subcondition. In order to determine a best-fitting dividing point, all the data were pooled (across condition and group) and pairs of lines were fitted to all contiguous combinations of signal durations. The best fit was achieved using durations of 2–10 ms and 20–200 ms for the two lines. This corresponds well with our hypothesized division between short-term and long-term temporal integration. However, it cannot be taken as providing strong support for this idea, since setting the division point at the adjacent shorter or longer durations only marginally worsened the goodness of fit. Nevertheless, for all further analysis, the division between durations less than or equal to 10 ms and those greater than 10 ms is maintained.

A series of *F*-tests showed that the improvement in fit using two lines was significant for the mean data in every condition [$F(2,29) > 3.49$, $p < 0.05$]. The resulting pairs of slopes for each listener and condition are given in Table III, together with the slopes for the mean data. Consider first the hearing-impaired group. In all but one case (MG, 40-dB condition), the slope for durations between 2 and 10 ms is steeper than for durations between 20 and 200 ms. Also, there seems to be no consistent effect of masker level. The mean data from the hearing-impaired listeners show no significant effect of masker level for either the short durations [$F(2,45) = 0.95$, $p > 0.3$] or the long durations [$F(2,33) = 1.96$, $p > 0.1$]. In contrast, for the normally hearing group, there is a strong effect of masker level at the short signal durations for all four individual listeners ($p \leq 0.01$), and for the mean data [$F(2,45) = 19.28$, $p \leq 0.0001$]. As can be seen from Table III, this is primarily due to the steeper slope of the 20-dB condition. Interestingly, however, differences in slope for the mean data at the longer durations are not significant [$F(2,33) = 2.17$, $p > 0.1$]. This is also true for the individual data for three of the four normally hearing listeners ($p > 0.1$); the marginally significant effect for the exception (listener MS; $0.01 < p < 0.05$) is due to the shallower slope of the 50-dB condition at the longer durations. Finally, an across-group comparison of the mean data revealed *no* significant difference in the shorter duration slopes between the hearing-impaired group, pooled across level, and the highest and lowest levels of the normally hearing group [$F(1,78) = 0.03$, $p > 0.5$]. The effect of group for the longer durations, however, was significant [$F(1,70) = 13.75$, p

TABLE III. Slopes from a two-line linear regression analysis [signal level (dB) against 10 log (duration)] of the data from Fig. 1. Masker spectrum level is in dB SPL. The first and second numbers denote the slopes for durations 2–10 ms and 20–200 ms, respectively. All slopes are negative, unless otherwise indicated.

| Group | Masker level | Listener | | | | Mean data |
|----------|--------------|------------|-------------|------------|------------|-------------------|
| | | AO | ST | DV | MS | |
| Normal | -10 dB | 0.99, 0.63 | 0.57, 0.40 | 1.22, 0.52 | 1.18, 0.60 | 0.99, 0.55 |
| | 20 dB | 1.86, 0.65 | 1.56, 0.63 | 1.70, 0.56 | 1.79, 0.65 | 1.72, 0.63 |
| | 50 dB | 0.95, 0.48 | 0.48, 0.33 | 0.86, 0.52 | 1.30, 0.26 | 0.89, 0.40 |
| Impaired | | MG | DT | VT | AW | |
| | 30 dB | 0.93, 0.25 | 1.34, 0.02 | 1.11, 0.40 | 0.70, 0.32 | 1.02, 0.24 |
| | 40 dB | 0.70, 0.84 | 0.85, 0.22 | 0.74, 0.46 | 1.04, 0.30 | 0.84, 0.46 |
| | 50 dB | 0.99, 0.38 | 0.80, +0.01 | 0.90, 0.44 | 1.03, 0.33 | 0.93, 0.28 |
| | | | | | | |

<0.001], indicating that the expected shallower slope for the hearing-impaired listeners is observed at the longer durations.

E. Discussion

The results confirm the main prediction made in the Introduction: For all four normally hearing listeners, the slope of the short-term integration function is steeper at the medium masker level than at the higher and lower levels. For the mean data, the slope is steeper by a factor of nearly 2. This is consistent with the idea that BM compression is greatest at medium sound levels. The lack of an effect of level for the hearing-impaired listeners is consistent with the idea that cochlear impairment leads to a linear BM response at all levels.

Surprisingly, the slopes of the short-term integration functions for hearing-impaired listeners at all three masker levels are very similar to those for the normally hearing listeners at the highest and lowest masker levels. If the slope of the short-term integration function is determined by peripheral compression, this implies that the BM response of the normally hearing listeners is approximately linear at very high and very low levels. This is consistent with some physiological data (e.g., Sellick *et al.*, 1982; Johnstone *et al.*, 1986; Ruggero and Rich, 1991) and is also consistent with the psychophysical results of Oxenham and Plack (1997). An alternative possibility is that the hearing-impaired listeners have some residual BM compression, which remains constant with level and is of the same order as the BM compression of normally hearing listeners at low and high levels. This seems unlikely, however, as the hearing losses of three of the four listeners between 6 and 8 kHz were between 50 and 60 dB; this is similar to the losses exhibited by the listeners in previous studies, where no residual compression was observed (Oxenham and Moore, 1995; Oxenham and Plack, 1997).

Equal integration slopes for normally hearing and hearing-impaired listeners at short durations have not been reported before. A review was therefore made of the available literature on temporal integration in hearing-impaired listeners at high frequencies (≥ 4 kHz) and short durations. The two most comparable studies are by Pedersen and Elberling (1973) and Florentine *et al.* (1988). Pedersen and Elberling (1973) measured temporal integration for durations between about 3.5 and 1000 ms at frequencies of, among others, 4 and 8 kHz. They found slopes, fitting all durations up to 200 ms, to be significantly shallower for their group of hearing-impaired listeners. Also, in the data of two sample subjects, it is clear that the integration function remains shallower even at the shortest durations. Furthermore, their use of relatively long onset and offset ramps makes it unlikely that the detection of splatter reduced the slope of the functions. The hearing-impaired listeners studied by Florentine *et al.* (1988) at 4 kHz show varied results for short durations. Defining the amount of temporal integration as the difference in threshold between a 2-ms and a 16-ms signal, four of the six hearing-impaired listeners show normal or near-normal integration, while the remaining two, listeners DP and PG, show reduced temporal integration. All the listeners in that

study had audiometric thresholds between 50 and 70 dB HL at 4 kHz, and there seems to be no correlation between amount of hearing loss at 4 kHz and the measured amount of short-term temporal integration.

Thus one previous study indicates that the reduction in temporal integration continues to very short signal durations (Pedersen and Elberling, 1973), while the majority of listeners in the other study produced slopes very similar to those of normally hearing listeners at short durations (Florentine *et al.*, 1988). One difference between both these studies and the present one is that our signals were presented in a gated noise, while in the other two studies signals were presented in quiet. It is not clear whether this difference could have affected the slope of the function.

Returning to the results from the normally hearing listeners, the change in the slope of the integration function with level implies that the signal-to-masker ratio changes with level for at least some signal durations. This has also been shown recently by von Klitzing and Kohlrausch (1994) for one listener. Using a 5-kHz signal with a total duration of 2 ms, they found that even when the signal was temporally centered in, or at the end of, a 300-ms noise masker, signal-to-masker ratios changed nonmonotonically with level by as much as 5 dB, reaching a maximum for an overall masker level of 60 dB SPL (20-dB spectrum level). In order to gain a better impression of how the integration function changes with level, thresholds were measured for signal durations of 2, 10, and 200 ms over a larger number of masker levels than were tested in experiment 1.

II. EXPERIMENT 2. CRITICAL RATIO AS A FUNCTION OF MASKER LEVEL: EFFECTS OF SIGNAL DURATION

A. Method

Thresholds were measured for half-amplitude signal durations of 2, 10, and 200 ms. For the normally hearing listeners, masker levels of 0-, 10-, 30-, and 40-dB spectrum level were tested. For the hearing-impaired listeners, levels of 35-, 45-, and 55-dB spectrum level were used. The stimuli, procedure, and listeners were all the same as those described in experiment 1.³

B. Results and discussion

The pattern of results was similar across the listeners in each group. For this reason, only the mean data are presented. The upper panel of Fig. 2 shows the data from the normally hearing listeners. Thresholds at -10-, 20-, and 50-dB masker levels are taken from experiment 1. Thresholds are plotted in terms of the ratio of the signal level to the noise spectrum level. Thus Weber's law would predict three parallel horizontal lines for the three conditions in the figure. However, consistent with the data of von Klitzing and Kohlrausch (1994), the signal-to-noise ratio (S/N_0 or SNR) for the shortest signal (open circles) increases at medium levels, reaching a maximum at 20- and 30-dB spectrum level. The maximum difference between levels is just over 4 dB, in good agreement with the 5 dB found by von Klitzing and Kohlrausch (1994). For the 10-ms signal (asterisks), the

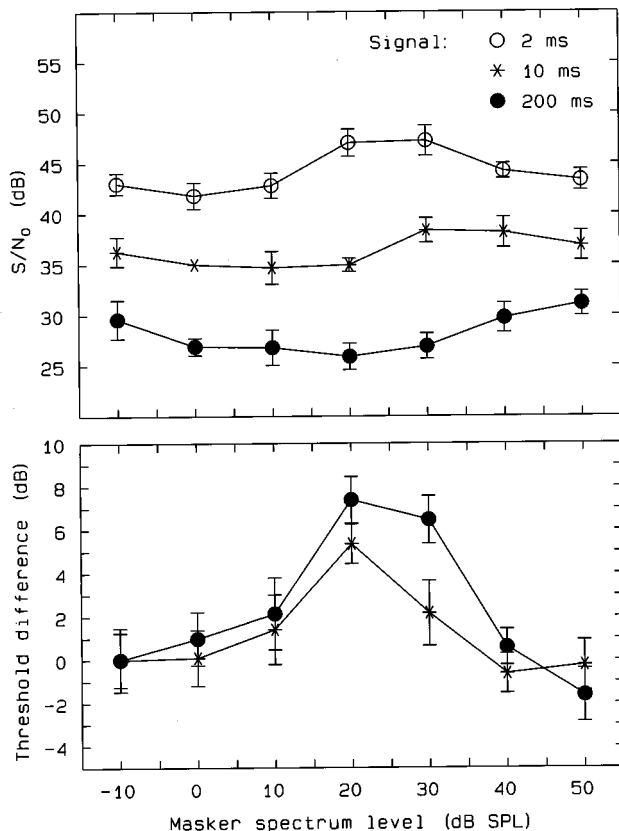


FIG. 2. Upper panel: Signal thresholds, in terms of signal-to-noise (spectrum level) ratio as a function of masker spectrum level, with signal duration as the parameter. Mean data of four listeners are shown, and error bars represent ± 1 standard error of the mean. Lower panel: Mean difference in thresholds between the 2- and 200-ms signal (filled circles) and between the 2- and 10-ms signal (asterisks), relative to the difference at -10 -dB spectrum level. Error bars represent ± 1 standard error of the mean.

variation with level is less systematic, while for the 200-ms signal (filled circles) the SNR is maximal at the highest and lowest levels, with a maximum variation of over 5 dB. The increase in SNR at the lowest level may be due to the approach to absolute threshold. The mean threshold in quiet for the long-duration signal was 12 dB SPL, and so the threshold in the presence of the -10 -dB noise was less than 8 dB above this level. The increase in SNR at the highest two levels may be due to the increase in the effective bandwidth of the auditory filter at high levels. The increase in the critical ratio with level for long-duration high-frequency signals in broadband noise has been found previously both in humans (Reed and Bilger, 1973; Moore, 1975; Pick, 1977) and in a behavioral study of masking in cats (Costalupes, 1983).

For each masker level, the difference in thresholds between the 2-ms and the 10-ms signal provides a rough measure of the amount of short-term integration. An estimate of overall integration can be obtained using the threshold difference between the 2-ms and the 200-ms signals. These differences, relative to the difference at the lowest masker level (-10 dB), are plotted in the lower panel of Fig. 2. It can be seen that the change in the difference with level is as much as 9 dB (20-dB vs 50-dB condition) and that, with the exception of the 30-dB condition, most of the change is due to the difference between the 2-ms and the 10-ms signals (as-

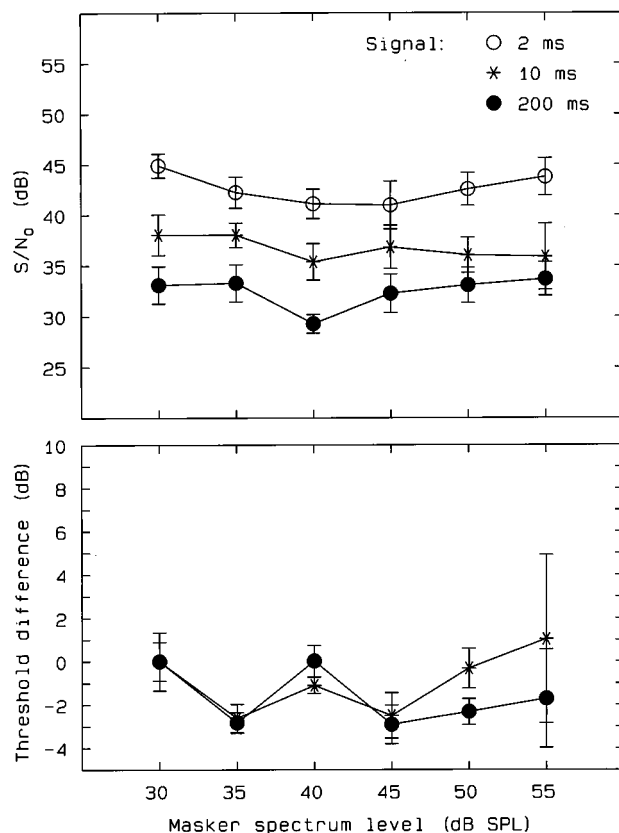


FIG. 3. As in Fig. 2, but for the four hearing-impaired listeners.

terisks). This corresponds to the finding in experiment 1, that there were significant differences in slope with level for short, but not for long, signal durations.

Data from the hearing-impaired listeners are shown in the upper panel of Fig. 3. Data for the 30-, 40-, and 50-dB conditions are taken from experiment 1. There is a tendency for the SNR of the 2-ms signal to be lower for masker levels in the middle of the range tested (40–45 dB spectrum level), and this is in part mirrored by the SNRs of the 200-ms signal. More importantly, however, there seems to be no systematic effect of level on the differences between the durations, as shown in the lower panel of Fig. 3. This is in agreement with the finding of experiment 1, that the slope of the integration function for the hearing-impaired listeners seems independent of masker level over the level range tested.

While the lack of a level effect for the hearing-impaired listeners is consistent with the expected changes in BM non-linearity, the strength of the conclusions is limited by the restricted ranges of levels (25 dB) over which the hearing-impaired listeners could be tested. Another caveat concerns the age difference between the two groups. It is possible that some of the difference in performance between the two groups reflects the large difference in mean ages, independent of hearing loss. While there is no *a priori* reason why age, independent of hearing loss, should affect the slope of the temporal-integration function, we cannot rule out that possibility based on our data. Finally, large intersubject variability has often been reported for hearing-impaired listeners

(e.g., Florentine *et al.*, 1988). While our listeners show reasonably consistent results, the possibility exists that other listeners with similar audiometric configurations may show a somewhat different pattern of results.

In summary, the results of experiment 2 confirm the strong mid-level effect for the normally hearing listeners and the lack of a level effect for the four hearing-impaired listeners over the testable range of levels. The following section is concerned with determining whether known changes in BM compression with level can provide a quantitative account of the data, at least at short signal durations.

III. MODELING THE EFFECTS OF PERIPHERAL COMPRESSION

In her influential paper, Penner (1978) derived the shape of a temporal-weighting function, $h(t)$, suitable for describing time-intensity trades, as:

$$h(t) = S_0 a p t^{ap-1}, \quad \text{for } t \geq 1,$$

$$h(t) = S_0, \quad \text{for } 0 < t < 1,$$

where t is time in arbitrary units, $-a$ is the slope of the integration function, measured in terms of signal level (dB SPL) against $10 \log [\text{duration (ms)}]$, p is the power to which signal intensity is raised, and S_0 is the initial height of the integrator. Clearly, ap must be less than unity. In practice, this is achieved, as it is assumed that $p < 1$ and the slope of the integration function ($-a$) in most cases lies between 0 and -1 . As the units of t can be made arbitrarily small, the discontinuity between $t = 1$ and $t < 1$ is of no practical importance. However, the equations show that for a given weighting function (from now on referred to as a temporal window), the slope of the integration function, $-a$, and the value of the power-law nonlinearity, p , are inversely proportional. An intuitive explanation of this relationship is given in the Appendix.

Within this framework, and assuming a fixed temporal window, the results from experiment 1 suggest that in order to account for the mid-level steepening in the slope of the short-term integration function, a nonlinearity is required which is about twice as compressive at medium sound levels as at high or low levels. Relating this to compression on the BM results in a prediction that, if the BM response of a damaged cochlea is linear, then the response of a normal cochlea is also linear at low and high levels and has a compressive growth, amounting to slightly more than 0.5 dB/dB at medium sound levels. This conclusion does not correspond well with the most recent physiological data. In most cases, compression resulting in growth of between 0.15 and 0.2 dB/dB at medium levels has been reported (e.g., Ruggero, 1992; Yates *et al.*, 1990; Murugasu and Russell, 1995). Also, other psychophysical experiments suggest that compression in human hearing is comparable to that measured physiologically (Oxenham and Moore, 1995; Oxenham and Plack, 1997). This difference may be related to the different stimuli used in the experiments. For instance, it is possible that the presence of broadband noise in the present experiment reduces the measured amount of compression. However, we know of no physiological measurements of the

overall BM response to broadband stimulation. Another reason for the apparent discrepancy may be found in one assumption of the Penner model, which requires closer consideration.

The assumption is that the output of the integrator is linearly related to the signal intensity raised to the power p , for a given signal duration. This condition is fulfilled when the signal is presented alone, as was assumed by Penner. However, in the case where a signal is presented simultaneously with a masker, the change in the output of the integrator due to the addition of the signal is no longer linearly related to the compressed signal intensity, I^p (for $p \neq 1$). Instead, the relationship between signal level and the integrator output due to the combination of signal and masker depends on the signal-to-noise ratio, on the amount of compression applied, and on the ratio of signal duration to temporal-window duration (assuming the masker is longer than the window). Hence, the equations derived by Penner (1978) are not valid for situations in which a masker and signal interact in the auditory periphery. In practice, this rules out Penner's model for all simultaneous-masking experiments.

A mathematical derivation of a revised analytical model, taking into account nonlinear interactions, is not attempted here. Instead, simulations were carried out using the model described below, in an attempt to find the change in nonlinearity with level necessary to account for our data within the context of the model.

A. Description of the model

The generic model we assume has been used many times in the past (e.g., Rodenburg, 1977; Viemeister, 1979; Buus and Florentine, 1985; Forrest and Green, 1987; Moore *et al.*, 1988; Plack and Moore, 1990) and consists of a bandpass-filter centered around the signal frequency (to simulate peripheral auditory filtering), a nonlinearity (rectification, followed by a power-law device), a short-term temporal integrator (or low-pass filter), and a decision device. The temporal window used here is a two-sided exponential window, as used by Oxenham and Moore (1994) to account for nonsimultaneous masking and by Peters *et al.* (1995) and Moore *et al.* (1996) to account for decrement detection. A symmetric shape is assumed, with the decay of the window on each side determined by a single time constant T , and given by the weighting function

$$W(t) = \exp(-|t|/T).$$

An asymmetric shape would be more realistic (Oxenham and Moore, 1994), but would not affect the outcome here. Data from temporal-integration experiments do not provide strong information as to the size of the temporal window. Therefore, we took the mean value of the "equivalent rectangular duration" (ERD), defined as $2T$, of 9.5 ms, from across a number of studies using stimulus frequencies of 4 kHz or greater (Oxenham and Moore, 1994; Peters *et al.*, 1995; Moore *et al.*, 1996).

The decision device used is also the same as, or similar to, that used in a number of previous studies (e.g., Plomp, 1964; Buus and Florentine, 1985; Oxenham and Moore, 1994). The output of the integrator due to the signal and

masker is compared with that due to the masker alone. If at any one time the difference between these two exceeds a criterion amount (in dB), the signal is “detected.” This criterion level provides the model with one free parameter.

Prior to compression, the stimuli within the model are represented simply by their envelopes, as it is thought that the auditory system has no access to stimulus fine structure at frequencies above about 4 kHz (Rose *et al.*, 1967). These envelopes are assumed to be flat for the steady-state portions of the stimuli and to have onset and offset ramps as used to gate the stimuli in the experiments. The noise masker and sinusoidal signal are assumed to add incoherently, i.e., the addition of two stimuli of equal level leads to a 3-dB increase in the overall level. The level of the masker envelope is initially derived by calculating the “effective” level of the broadband-noise masker within the equivalent rectangular bandwidth (ERB) of an auditory filter centered around 6.5 kHz (Glasberg and Moore, 1990). However, for the data from the hearing-impaired listeners, and for the 50-dB data from the normally hearing listeners, this value was increased by 2 dB to take account of the presumed broadening of the auditory filters in these conditions; a 2-dB increase corresponds to a broadening by a factor of about 1.6. The assumed masker level for the -10 -dB condition for the normally hearing listeners was also increased by 2 dB. This was done to model the effect of an internal noise, which is assumed to be added independently and to be responsible for absolute threshold. Thus only the 20-dB condition for the normally hearing listeners retained the original effective masker level; all other conditions were simulated using a masker level 2 dB higher. Any smoothing of the envelope due to the auditory filter at 6.5 kHz is assumed to be negligible compared to the smoothing of the temporal window.

In using a flat temporal envelope to represent the masker, we ignore the noise’s variability in level, as well as its envelope distribution. Regarding the first point, as long as a fixed temporal integrator is assumed (as we do here), the slope of the predicted integration function is the same whether or not the random level fluctuations of the noise are taken into account (Eddins and Green, 1995). The second point concerns the effective (long-term) level of a time-varying stimulus once it has been compressed. Two stimuli of equal energy, one with a flat and the other with a modulated temporal envelope, will have different mean levels after being compressed: The time-varying stimuli will have a lower mean level, as the peaks of the envelope will be compressed. This is also true when comparing a sinusoid with a Gaussian noise. However, initial simulations taking into account the envelope distribution of a Gaussian noise and the envelope distribution for a sinusoidal signal in Gaussian noise (van de Par and Kohlrausch, 1995) showed very little difference between this and using a flat temporal envelope to represent the noise. Using “realistic” signal-to-noise ratios and compression values, the maximum difference in absolute predictions reached 0.8 dB, and the maximum difference in the predicted slope of the integration function was 0.5%. Thus, for simplicity, in all the simulations presented below, the noise was represented by a flat temporal envelope.

Finally, possible dynamic (time-variant) effects in the

auditory system, such as the large onset response in the auditory nerve, are not taken into account here. While BM compression is thought to be near instantaneous (and hence time-invariant), the rate-intensity functions of auditory nerve fibers are dependent on prior levels of adaptation. Some previous psychoacoustic models have incorporated an approximation to this response (Zwislocki, 1969; Dau *et al.*, 1996a), although its importance in perception is not well understood. Nevertheless, the *change* in overall response due to changes in BM compression may not be affected by the response at the level of the auditory nerve. Thus the exclusion of this aspect of auditory processing from the model may not affect the main conclusions.

B. Model predictions

In order to derive the best-fitting criterion parameter for the model, the group mean data from experiment 1 for durations up to 10 ms were used from all conditions except those from the 20-dB (normally hearing) condition. Recall that the slopes of these data were not significantly different from each other. We fitted these data with the model by assuming that the signals are processed linearly before being integrated. “Linear” in these terms is with respect to intensity, rather than amplitude, for two reasons. First, Oxenham and Moore (1995) found that for hearing-impaired listeners, data from the additivity of nonsimultaneous masking could be accounted for well by assuming linear additivity of intensity. Second, physiological studies by Yates and colleagues (Yates *et al.*, 1990; Yates, 1990) have found that, in the absence of BM compression, the rate-intensity function of auditory-nerve fibers is a linear function of stimulus intensity.

By simulating all the conditions, and by comparing the predictions with the mean data from the experiment, the best-fitting (least-squared error) criterion value was selected. The mean data from experiment 1 for durations between 2 and 10 ms are replotted in Fig. 4, together with the model predictions (solid curves). The best-fitting decision criterion was equivalent to a steady-state level difference of 5.6 dB.

Next, the same model was used to fit the data from the remaining 20-dB condition. Here, the value of the nonlinearity was varied to produce the best-fitting predictions, while the time constant and the decision criterion were held constant. The decision criterion was set in terms of an equivalent long-duration (steady-state) level difference (in dB) prior to compression. The level difference prior to compression, rather than the “internal” (compressed) level difference, was chosen in order to maintain an approximation to Weber’s law for long-duration stimuli. This was done for empirical reasons and implies more efficient coding at medium levels than at high or low levels. As discussed above, the model of Penner (1978) predicts a best-fitting exponent of about 0.54, as the slope of the 20-dB condition is about 1.84 (the reciprocal) times steeper than that of the other conditions. For the present model, however, the best-fitting exponent was 0.25, indicating a compression ratio of 4:1. This value is more in line with the physiological estimates of BM nonlinearity, as mentioned above. Overall, the fit is very good, and lies within one standard error of the mean for 27 out of 30 data

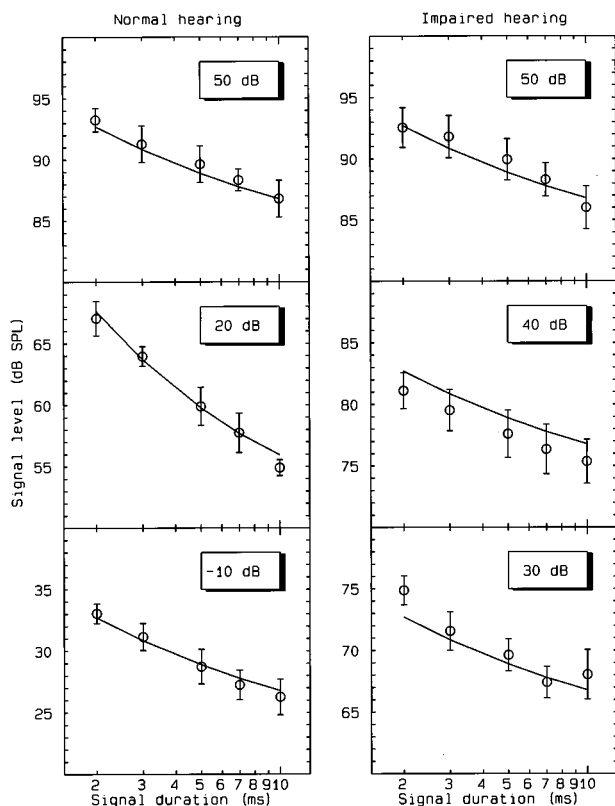


FIG. 4. Mean data from Fig. 1 re-plotted for signal durations between 2 and 10 ms (symbols) together with the predictions of the model described in the text (solid curves). Error bars represent ± 1 standard error of the mean.

points. The predictions for the 40-dB condition of the hearing-impaired listeners lie consistently between 1 and 1.5 dB above the data points. This reflects the fact that, for our four listeners, the critical ratio in this condition seems to be lower than in the other conditions. If we allowed the decision criterion to vary across levels, the fit would improve.

As stated in the Introduction, longer-term integration, for signal durations of more than 10 ms, may be due to longer time constants (in parallel with, or following, our hypothesized short-term integrator), a multiple-looks mechanism, an analysis window of variable duration, or a combination of these. At present, there seems to be no good way of distinguishing between these possibilities. Similarly, if we accept that the BM response in the normal cochlea is completely linear at low levels, then the reason for the reduced temporal integration at longer durations in the hearing-impaired listeners remains unclear. One of this study's initial aims of accounting for differences between normal and abnormal temporal integration therefore remains partially unfulfilled.

One hypothesis for explaining the difference in temporal integration at longer durations has been proposed by Carlyon *et al.* (1990). They found that the psychometric functions for single 5-ms 1-kHz tone pulses were steeper for hearing-impaired listeners than for normally hearing listeners. Psychometric functions were also steeper for a series of ten such pulses, separated by 80 ms. Thus while temporal integration, measured in terms of the level difference at threshold be-

tween one and ten pulses, was reduced in the hearing-impaired listeners, the change in detectability, in terms of d' , was the same as for normally hearing listeners. It may be, therefore, that a change in the underlying psychometric function can account for changes in the slope of the longer duration temporal-integration function, although this has not been tested for tones of different durations.

C. Concluding remarks

The model presented here provides a good description of short-term temporal integration, and shows that a change in nonlinearity consistent with that found in physiological studies can be applied to account for changes in temporal integration with level. However, the data are not suitable for deriving precise estimates of the weighting function for the temporal integrator. In this study we used as a time constant the mean value derived from other studies, but even a doubling of the value of this time constant had only a small effect on the mean squared error of the predictions. This is because the data we fitted only extend to durations of 10 ms, and increases in the ERD beyond values of about 10 ms have an increasingly small effect on predictions. Furthermore, the quality of the predictions is not dependent on the exact form of the temporal window. A 10-ms rectangular window, a Hanning window with a total duration of 20 ms, and a low-pass filter with a cutoff frequency of about 50 Hz all provide reasonably good fits to the data. In all these cases, the best fit to the data from the 20-dB condition was achieved with a compressive exponent of between 0.2 and 0.3. In general, a longer-duration temporal window results in a smaller required change in the nonlinearity for a given change in slope.

IV. SUMMARY

(1) For a signal frequency of 6.5 kHz, the slope of the short-term temporal-integration function for normally hearing listeners is steeper at medium levels than at high or low levels by a factor of nearly 2. This is consistent with the hypothesis that BM compression is greatest at medium levels, between about 50 and 70 dB SPL. For the hearing-impaired listeners, no effect of level was found over the 25-dB range tested. This is consistent with the idea that cochlear damage can lead to a linear BM input-output function.

(2) For signal durations of 10 ms and less, there was no significant difference in slope between the data from the hearing-impaired listeners and those from the normally hearing listeners at the low and high masker levels. While this is inconsistent with the results of one previous study (Pedersen and Elberling, 1973), four of six hearing-impaired listeners in a later study (Florentine *et al.*, 1988) show results similar to ours. If future studies confirm this finding, it suggests that the normal BM input-output function at low and high levels may be similar to that of listeners with cochlear hearing loss and so may be approximately linear.

(3) For signal durations of 20 ms and more, the temporal-integration functions for the hearing-impaired listeners were generally shallower than for the normally hear-

ing listeners. This is in agreement with the literature. However, the mechanisms underlying this effect remain unclear.

(4) The model proposed by Penner (1978), which predicts that the slope of integration is inversely proportional to the amount of compression for a given temporal window, is shown not to be applicable to any simultaneous-masking conditions.

(5) Simulations using an established model of temporal resolution indicate that the change in compression with level necessary for the model to account for the data from normally hearing listeners is similar to that found in many physiological measurements of BM compression and other psychophysical tasks.

ACKNOWLEDGMENTS

The work was supported by the MRC and by a Wellcome Trust Travelling Research Fellowship (0044215/Z/95/Z) awarded to AJO. We thank Armin Kohlrausch, Reinier Kortekaas, Steven van de Par, and the reviewers, Torsten Dau and Mary Florentine, for helpful comments on previous versions of the manuscript.

APPENDIX: COMPRESSION AND INTEGRATION

Here, an intuitive explanation is given for why compression prior to integration produces a steeper integration function. Assume a rectangular temporal-weighting function (integrator) of longer duration than the signal, operating on stimulus intensity (I^p , where $p=1$), and consider a signal at threshold. If the signal duration is doubled, the signal intensity must be halved to maintain the same (threshold) output of the integrator. This corresponds to a decrease in level of 3 dB. If instead the rectified signal amplitude is integrated, when the duration is doubled, the *amplitude* of the signal must be halved, leading to a decrease in level of 6 dB for the same change in duration. Since the amplitude is proportional to a compressed version of the intensity (I^p , where $p=0.5$), compression produces a steeper slope of the integration function for a given temporal window. In these terms a halving of the exponent produces a doubling of the slope of the function.

¹One listener (MS) reported that he found the loudness of the 50-dB masker uncomfortable. For this listener, the spectrum level of the masker was reduced to 45 dB before data were collected.

²In this study, the term hearing impaired is used for listeners with absolute thresholds higher than 20 dB HL. If age is taken into account, however, the hearing of listener AW falls within the "normal" range for men aged 80 and above (Morrell *et al.*, 1996).

³For the normally hearing listeners, thresholds for the 10-ms signal were measured at a later time than the other thresholds. At that time, listeners AO and ST were tested on different equipment. However, replications of thresholds from other conditions revealed no consistent differences, with deviations no greater than 1.2 dB.

Buus, S., and Florentine M. (1985). "Gap detection in normal and impaired listeners: The effect of level and frequency," in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, New York), pp. 159–179.

Carlyon, R. P., Buus, S., and Florentine, M. (1990). "Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners," *J. Acoust. Soc. Am.* **87**, 260–268.

Chung, D. Y. (1981). "Masking, temporal integration, and sensorineural hearing loss," *J. Speech Hear. Res.* **24**, 514–520.

Costalupes, J. A. (1983). "Broadband masking noise and behavioral pure tone thresholds in cats," *J. Acoust. Soc. Am.* **74**, 758–764.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* (submitted).

Dau, T., Püschel, D., and Kohlrausch, A. (1996a). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Am.* **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (1996b). "A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements," *J. Acoust. Soc. Am.* **99**, 3623–3631.

Eddins, D. A., and Green, D. M. (1995). "Temporal integration and temporal resolution," in *Hearing*, edited by B. C. J. Moore (Academic, London), Vol. 6 of *Handbook of Perception and Cognition*.

Elliott, L. L. (1963). "Tonal thresholds for short-duration stimuli as related to subject hearing level," *J. Acoust. Soc. Am.* **35**, 578–583.

Feldtkeller, R., and Oetinger, R. (1956). "Die Hörbarkeitsgrenzen von Impulsen verschiedener Dauer," *Acustica* **6**, 489–493.

Florentine, M., Buus, S., and Poulsen, T. (1996). "Temporal integration of loudness as a function of level," *J. Acoust. Soc. Am.* **99**, 1633–1644.

Florentine, M., Fastl, H., and Buus, S. (1988). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," *J. Acoust. Soc. Am.* **84**, 195–203.

Forrest, T. G., and Green, D. M. (1987). "Detection of partially filled gaps in noise and the temporal modulation transfer function," *J. Acoust. Soc. Am.* **82**, 1933–1943.

Garner, W. R., and Miller, G. A. (1947). "The masked threshold of pure tones as a function of duration," *J. Exp. Psychol.* **37**, 293–303.

Gengel, R. W. (1972). "Auditory temporal integration at relatively high masked-threshold levels," *J. Acoust. Soc. Am.* **51**, 1849–1851.

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.* **47**, 103–138.

Green, D. M., Birdsall, T. G., and Tanner, W. P. (1957). "Signal detection as a function of signal intensity and duration," *J. Acoust. Soc. Am.* **29**, 523–531.

Hughes, J. W. (1946). "The threshold of audition for short periods of stimulation," *Proc. R. Soc. London, Ser. B* **133**, 486–490.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (1986). "Basilar membrane measurements and the travelling wave," *Hearing Res.* **22**, 147–153.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.

Miskolczy-Fodor, F. (1953). "Monaural loudness-balance test and determination of recruitment degree with short sound-impulses," *Acta Oto-Laryngol.* **43**, 573–595.

Moore, B. C. J. (1975). "Mechanisms of masking," *J. Acoust. Soc. Am.* **57**, 391–399.

Moore, B. C. J. (1991). "Characterization and simulation of impaired hearing: Implications for hearing aid design," *Ear Hear.* **12**, Suppl. 154S–161S.

Moore, B. C. J. (1995). *Perceptual Consequences of Cochlear Damage* (Oxford U.P., Oxford, England).

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (1988). "The shape of the ear's temporal window," *J. Acoust. Soc. Am.* **83**, 1102–1116.

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.

Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (1996). "Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level," *J. Acoust. Soc. Am.* **100**, 1949–1967.

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.

Oxenham, A. J., and Moore, B. C. J. (1994). "Modeling the additivity of nonsimultaneous masking," *Hearing Res.* **80**, 105–118.

Oxenham, A. J., and Moore, B. C. J. (1995). "Additivity of masking in normally hearing and hearing-impaired subjects," *J. Acoust. Soc. Am.* **98**, 1921–1934.

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**, ■■■–■■■.
- Pedersen, C. B., and Elberling, C. (1973). "Temporal integration of acoustic energy in patients with presbycusis," Acta Oto-Laryngol. **75**, 32–37.
- Penner, M. J. (1978). "A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts," J. Acoust. Soc. Am. **63**, 195–201.
- Penner, M. J., and Cudahy, E. (1973). "Critical masking interval: A temporal analog of the critical band," J. Acoust. Soc. Am. **54**, 1530–1534.
- Penner, M. J., Robinson, C. E., and Green, D. M. (1972). "The critical masking interval," J. Acoust. Soc. Am. **52**, 1661–1668.
- Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (1995). "Effects of level and frequency on the detection of decrements and increments in sinusoids," J. Acoust. Soc. Am. **97**, 3791–3799.
- Pick, G. (1977). "Comment on Scharf and Meiselman," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 233–234.
- Plack, C. J., and Moore, B. C. J. (1990). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187.
- Plomp, R. (1964). "The rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**, 277–282.
- Plomp, R., and Bouman, M. A. (1959). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**, 749–758.
- Reed, C. M., and Bilger, R. C. (1973). "A comparative study of *S/No* and *E/No*," J. Acoust. Soc. Am. **53**, 1039–1044.
- 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.
- Rodenburg, M. (1977). "Investigation of temporal effects with amplitude modulated signals," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
- Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–793.
- Ruggero, M. A. (1992). "Responses to sound of the basilar membrane of the mammalian cochlea," Curr. Op. Neurobiol. **2**, 449–456.
- 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., Robles, L., and Recio, A. (1995). "The effects of acoustic overstimulation, other cochlear injury and death on basilar membrane responses to sound," in *Effects of Noise on Hearing: Vth International Symposium*, edited by R. J. Salvi, A. Axelsson, D. Henderson, and R. Hamernik (Thieme Medical, Stockholm).
- Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," J. Acoust. Soc. Am. **72**, 131–141.
- Stephens, S. D. G. (1973). "Auditory temporal integration as a function of intensity," J. Sound Vib. **37**, 235–246.
- van de Par, S., and Kohlrausch, A. (1995). "Analytical expressions for the envelope correlation of certain narrow-band stimuli," J. Acoust. Soc. Am. **98**, 3157–3169.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
- 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.
- Wright, H. N. (1968). "Clinical measurement of temporal auditory summation," J. Speech Hear. Res. **11**, 109–127.
- Yates, G. K. (1990). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions," Hearing Res. **50**, 145–162.
- Yates, G. K., Winter, I. M., and Robertson, D. (1990). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hearing Res. **45**, 203–220.
- Zwicker, E. (1960). "Ein Verfahren zur Berechnung der Lautstärke," Acustica **10**, 304–308.
- Zwislocki, J. J. (1960). "Theory of temporal auditory summation," J. Acoust. Soc. Am. **32**, 1046–1060.
- Zwislocki, J. J. (1969). "Temporal summation of loudness: An analysis," J. Acoust. Soc. Am. **46**, 431–441.

Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration^{a)}

Jungmee Lee^{b)} and Sid P. Bacon

Department of Speech and Hearing Science, P.O. Box 871908, Arizona State University, Tempe, Arizona 85287-1908

(Received 1 July 1996; accepted for publication 9 January 1997)

Discrimination of the change in depth of sinusoidal amplitude modulation (AM) was investigated as a function of stimulus duration. The carrier frequency was 4000 Hz, the standard modulation depth (m) was either 0.1, 0.18, or 0.3, and the modulation rate was either 10, 20, 40, or 80 Hz. For all standard depths and modulation rates, threshold (Δm) decreased by more than a factor of two as stimulus duration doubled from the shortest duration used up to a certain duration (critical duration), beyond which the threshold decreased only slightly or remained constant. The critical duration corresponded to about four cycles of modulation. Psychometric functions were measured for different stimulus durations to examine the extent to which a multiple-looks model could explain the present data. This model provided a reasonable prediction of the change in AM depth discrimination threshold as a function of stimulus duration. © 1997 Acoustical Society of America. [S0001-4966(97)01206-X]

PACS numbers: 43.66.Mk, 43.66.Ba [JWH]

INTRODUCTION

This study investigated the effect of stimulus duration on amplitude modulation (AM) depth discrimination. A general characteristic of auditory behavior is that detection and discrimination improve with increasing stimulus duration. This improvement suggests the operation of a temporal-integration process. Previous studies have measured the effect of stimulus duration on the detection of beats or AM (Viemeister, 1970, 1979; Sheft and Yost, 1990) and on the discrimination of AM rate (Lee, 1994). In general, they showed that performance improved with increasing stimulus duration, although the results differed with regard to the function relating performance to stimulus duration.

Viemeister (1970) investigated the effect of stimulus duration on beat detection. He found that at a beat frequency of 4 Hz, detectability improved with increasing stimulus duration from 125 to 500 ms. The detectability increased by roughly the square root of the number of beats. In 1979, Viemeister examined the effect of stimulus duration on AM detection, using stimulus durations of 250, 500, and 1500 ms. More recently, Sheft and Yost (1990) studied AM detection for durations from 400 ms to a duration corresponding to one cycle of modulation. Both studies showed that the detection threshold decreased gradually with increasing stimulus duration.

For AM rate discrimination, however, Lee (1994) found that at modulation rates of 20 and 40 Hz, the discrimination threshold decreased abruptly (more than a factor of two) as the stimulus duration increased to a certain duration (critical duration), beyond which threshold decreased slightly or re-

mained constant. The critical duration corresponded to about five cycles of modulation.

To date, no one has studied the effect of stimulus duration on AM depth discrimination. The function relating AM depth discrimination to stimulus duration might be predicted to be more like that for AM detection, because the critical variable for both tasks is related to modulation depth. On the other hand, the function for AM depth discrimination might be more like that for AM rate discrimination, because both are related to envelope discrimination. It will be worthwhile, therefore, to investigate how AM depth discrimination is affected by stimulus duration.

I. EXPERIMENT 1: DISCRIMINATION OF AM DEPTH AS A FUNCTION OF DURATION

A. Method

1. Subjects

Three normal-hearing subjects participated in this study. They ranged in age from 23 to 30 years. Each subject's absolute thresholds were not worse than 10 dB HL (ANSI, 1989) at any of the octave frequencies from 0.25 to 8 kHz. All subjects had experience in other psychoacoustic experiments. One of the subjects (S1) was author JL; the other two subjects were paid an hourly wage for their participation.

2. Stimuli

The stimuli throughout the experiments are as follows

$$s(t) = A \{ 1 + [(m + \Delta m) \cos(2\pi f_m t + \phi)] \} \times [\cos(2\pi f_c t)], \quad (1)$$

where m is the standard modulation depth, Δm is the change in the modulation depth, f_m is the modulation rate, ϕ is the phase of modulation, and f_c is the carrier frequency.

^{a)}This research was presented at the 131st meeting of the Acoustical Society of America [J. Lee and S. P. Bacon, "Amplitude modulation depth discrimination of a sinusoidal carrier," J. Acoust. Soc. Am. **99**, 2566(A) (1996)].

^{b)}Corresponding author: Electronic mail: jlee@asu.edu

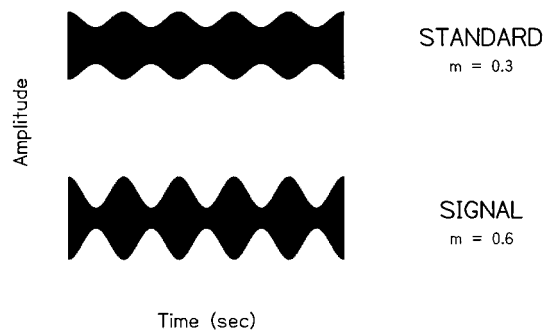


FIG. 1. An example of the standard (top) and signal (bottom) stimulus.

The stimuli were generated digitally and produced (TDT DA1) at a 20-kHz sampling rate. The output of the 16-bit digital-to-analog converter was low-pass filtered at 8 kHz (Kemo VBF25.01, 135 dB/oct). The onsets and offsets of all stimuli were shaped by a cosine-squared function giving a rise–fall time of 5 ms. The duration of the stimulus, including rise–fall time, was either 25, 50, 100, 200, 400, or 800 ms. The modulation rate was 10, 20, 40, or 80 Hz. The standard modulation depth (m) was 0.1, 0.18, or 0.3. The carrier frequency was 4000 Hz. The overall level of the carrier was fixed at 60 dB SPL.¹ The phase of modulation was randomly chosen between trials.

3. Procedure

A two-interval, forced-choice paradigm was used to measure the threshold (Δm) for discriminating modulation depth. The psychophysical task was to discriminate a “standard” stimulus having a modulation depth of m , from a “signal” stimulus having modulation depth of $m + \Delta m$, by identifying the interval having the greater modulation depth. Figure 1 shows an example of the standard and the signal stimulus. The modulation depth of the signal was varied in steps of 1 dB [in $20 \log(m + \Delta m)$] using a three-down, one-up procedure, estimating the 79.4% correct point on the psychometric function (Levitt, 1971). The two observation intervals were separated by 500 ms. Each run consisted of 60 trials. The first two or three reversals were discarded (two if the total number of reversals was even, three if it was odd) and the remaining reversals were averaged to obtain the threshold estimate for that run. Threshold estimates were discarded on the rare occasions when the standard deviation of the reversals was greater than 5 dB or when there were fewer than six reversals in the mean. Each threshold reported is the mean of the estimates from at least three runs. When the standard deviation of the three estimates exceeded 3 dB, an additional estimate was obtained, and all estimates were averaged. This continued until the standard deviation of the estimates was less than 3 dB or until six estimates were obtained. All of the thresholds reported here had standard deviations less than 3 dB.

Subjects listened monaurally through a TDH-49P headphone while seated in a soundproof room and responded by pushing one of two buttons on a response panel. Lights were used to indicate the observation intervals and then to provide correct-response feedback.

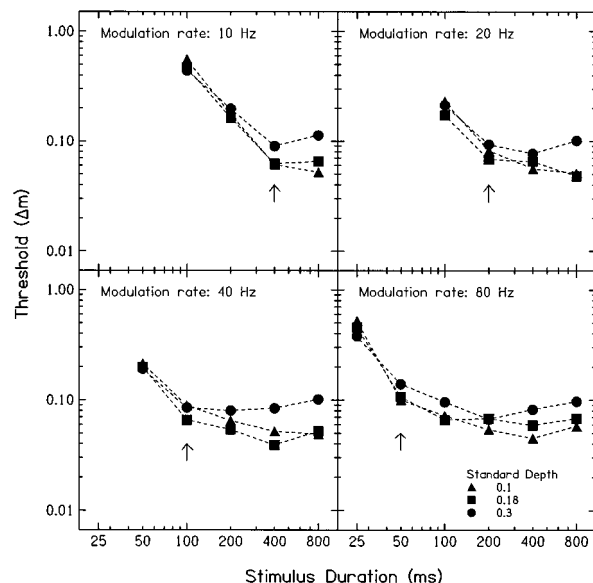


FIG. 2. Discrimination threshold Δm as a function of stimulus duration at modulation rates of 10, 20, 40, and 80 Hz, results are the average of three subjects. Three standard modulation depth were used: 0.1 (triangles), 0.18 (squares), and 0.3 (circles). The arrows indicate the duration corresponding to four cycles of modulation.

B. Results

Figure 2 shows the discrimination threshold, Δm , as a function of the stimulus duration at modulation rates of 10, 20, 40, and 80 Hz. The different symbols in each panel represent the different standard modulation depths: 0.1 (triangles), 0.18 (squares), and 0.3 (circles). Only mean results are presented in this figure. Individual data are presented in Table I.

For each modulation rate, the discrimination threshold decreases with increasing stimulus duration, regardless of the standard modulation depth. The threshold decreases by more than a factor of two as stimulus duration is doubled from the shortest duration used up to a certain duration (critical duration), which is indicated by an arrow in Fig. 2. The critical duration corresponds to four cycles of modulation for each modulation rate. For the stimulus durations beyond the critical duration, the threshold decreases only slightly if at all.

The results for each modulation rate are analyzed with a two-factor (stimulus duration by standard modulation depth) repeated measures ANOVA. The analysis indicates that the effect of stimulus duration is significant (for a 10-Hz rate, $F_{(3,24)} = 120.343$, $p < 0.001$; for a 20-Hz rate, $F_{(3,24)} = 37.163$, $p < 0.001$; for a 40-Hz rate, $F_{(4,30)} = 28.810$, $p < 0.001$; for a 80-Hz rate, $F_{(5,36)} = 97.121$, $p < 0.001$), but that the effect of standard modulation depth is not (for a 10-Hz rate, $F_{(2,24)} = 1.159$, $p > 0.1$; for a 20-Hz rate, $F_{(2,24)} = 2.907$, $p > 0.1$; for a 40-Hz rate, $F_{(2,30)} = 2.253$, $p > 0.1$; for an 80-Hz rate, $F_{(2,36)} = 0.079$, $p > 0.1$). In addition, the interaction between stimulus duration and standard modulation depth is not significant (for a 10-Hz rate, $F_{(6,24)} = 1.814$, $p > 0.1$; for a 20-Hz rate, $F_{(6,24)} = 0.879$, $p > 0.1$; for a 40-Hz rate, $F_{(8,30)} = 0.651$, $p > 0.1$; for an 80-Hz rate, $F_{(10,36)} = 1.813$, $p > 0.05$).

TABLE I. Individual and mean thresholds (Δm) for a standard modulation depth (m) of 0.1, 0.18, and 0.3. Threshold was measured for different modulation rates at different stimulus durations.

| | | 0.1 | | | | 0.18 | | | | 0.3 | | | |
|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | S1 | S2 | S3 | Mean | S1 | S2 | S3 | Mean | S1 | S2 | S3 | Mean |
| 10 Hz | 100 ms | 0.475 | 0.580 | 0.613 | 0.556 | 0.402 | 0.469 | 0.540 | 0.470 | 0.467 | 0.409 | 0.447 | 0.441 |
| | 200 ms | 0.149 | 0.246 | 0.132 | 0.176 | 0.128 | 0.269 | 0.094 | 0.164 | 0.169 | 0.293 | 0.132 | 0.198 |
| | 400 ms | 0.038 | 0.096 | 0.050 | 0.061 | 0.043 | 0.105 | 0.037 | 0.062 | 0.093 | 0.117 | 0.060 | 0.090 |
| | 800 ms | 0.029 | 0.077 | 0.049 | 0.052 | 0.049 | 0.108 | 0.037 | 0.065 | 0.110 | 0.146 | 0.082 | 0.113 |
| 20 Hz | 100 ms | 0.177 | 0.256 | 0.259 | 0.231 | 0.122 | 0.248 | 0.148 | 0.173 | 0.195 | 0.273 | 0.175 | 0.214 |
| | 200 ms | 0.069 | 0.100 | 0.077 | 0.082 | 0.067 | 0.057 | 0.082 | 0.068 | 0.067 | 0.097 | 0.114 | 0.093 |
| | 400 ms | 0.031 | 0.079 | 0.058 | 0.056 | 0.047 | 0.086 | 0.062 | 0.065 | 0.074 | 0.072 | 0.086 | 0.077 |
| | 800 ms | 0.027 | 0.080 | 0.046 | 0.051 | 0.047 | 0.057 | 0.041 | 0.048 | 0.088 | 0.075 | 0.141 | 0.101 |
| 40 Hz | 50 ms | 0.152 | 0.244 | 0.246 | 0.214 | 0.157 | 0.257 | 0.182 | 0.199 | 0.162 | 0.234 | 0.176 | 0.191 |
| | 100 ms | 0.068 | 0.086 | 0.112 | 0.089 | 0.045 | 0.065 | 0.089 | 0.066 | 0.067 | 0.108 | 0.080 | 0.085 |
| | 200 ms | 0.054 | 0.066 | 0.075 | 0.065 | 0.043 | 0.054 | 0.065 | 0.054 | 0.056 | 0.124 | 0.059 | 0.080 |
| | 400 ms | 0.030 | 0.062 | 0.062 | 0.052 | 0.033 | 0.038 | 0.046 | 0.039 | 0.059 | 0.137 | 0.055 | 0.084 |
| | 800 ms | 0.036 | 0.054 | 0.057 | 0.049 | 0.044 | 0.072 | 0.039 | 0.052 | 0.052 | 0.170 | 0.080 | 0.101 |
| 80 Hz | 25 ms | 0.500 | 0.665 | 0.402 | 0.522 | 0.458 | 0.399 | 0.511 | 0.456 | 0.394 | 0.378 | 0.364 | 0.379 |
| | 50 ms | 0.073 | 0.111 | 0.118 | 0.100 | 0.058 | 0.130 | 0.132 | 0.107 | 0.112 | 0.150 | 0.156 | 0.139 |
| | 100 ms | 0.040 | 0.076 | 0.099 | 0.072 | 0.040 | 0.085 | 0.074 | 0.066 | 0.064 | 0.127 | 0.096 | 0.096 |
| | 200 ms | 0.026 | 0.061 | 0.076 | 0.054 | 0.037 | 0.121 | 0.044 | 0.068 | 0.061 | 0.099 | 0.040 | 0.067 |
| | 400 ms | 0.017 | 0.060 | 0.057 | 0.045 | 0.033 | 0.096 | 0.050 | 0.059 | 0.056 | 0.142 | 0.049 | 0.082 |
| | 800 ms | 0.050 | 0.060 | 0.063 | 0.057 | 0.037 | 0.120 | 0.048 | 0.043 | 0.063 | 0.160 | 0.069 | 0.066 |

C. Discussion

Consider first the results for longer durations (400 and 800 ms). For all modulation rates, discrimination thresholds were between about 0.04 and 0.07 for the standard modulation depth of 0.1 and 0.18, and between 0.07 and 0.1 for the standard depth of 0.3. These values are consistent with those in other studies (Grantham and Bacon, 1988; Ozimek and Sek, 1988; Wakefield and Viemeister, 1990; Moore *et al.*, 1991).

The functions relating discrimination threshold to stimulus duration are consistent with those for AM rate discrimination (Lee, 1994), where the critical duration corresponded to about five cycles of modulation. The present results, however, are not consistent with those of AM detection (Viemeister, 1979; Sheft and Yost, 1990), where it has been shown that thresholds decrease gradually with increasing stimulus duration, without showing a clear breakpoint at the duration corresponding to four or five cycles of modulation. Viemeister's (1970) results on the effect of stimulus duration on beat detection did not show a breakpoint either, although his range of stimulus duration resulted in only one-half to two beats per observation interval.

The difference between the effect of stimulus duration on AM detection and AM discrimination is shown in Fig. 3, where the AM depth discrimination thresholds from the present study (averaged across the three standard depths) are plotted together with the AM detection thresholds from Sheft and Yost (1990; averaged across their two subjects). The discrimination thresholds (unfilled circles) are referred to the left-hand axis, whereas the detection thresholds (filled circles) are referred to the right-hand axis. Both types of thresholds are plotted in decibels, as $20 \log \Delta m$ or $20 \log m$. For all modulation rates, the detection and discrimination thresholds show similar trends at stimulus dura-

tions yielding four or more cycles of modulation (the arrows in each panel indicate the duration corresponding to four cycles). The functions diverge, however, at shorter durations.

One possible reason for the separation of the detection and discrimination functions at the shorter durations may be related to the inherent difficulty of the two tasks. For modulation detection, the subject needs only to determine which of two stimuli was modulated. For the discrimination task, on the other hand, the subject must determine which of two

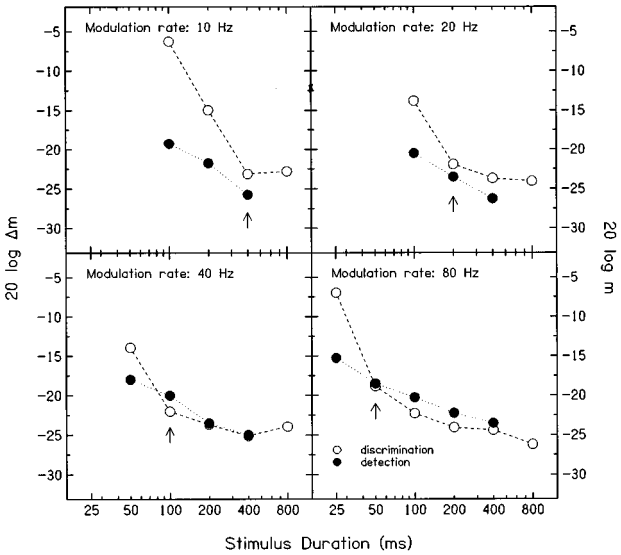


FIG. 3. Comparison of the effect of duration on AM detection with that on AM depth discrimination. The AM detection thresholds (filled symbols) are from Sheft and Yost (1990; averaged across their two subjects), and are referred to the right-hand axis. The AM depth discrimination thresholds (unfilled symbols) are from the present study (averaged across the three standard depths), and are referred to the left-hand axis. The arrows indicate the duration corresponding to four cycles of modulation.

stimuli had the greater modulation depth. This could conceivably require a better internal representation of the stimuli, and this representation might be adversely affected at the shorter durations yielding a small number of modulation cycles. As pointed out by an anonymous reviewer, this line of reasoning implies that AM depth discrimination at short durations should improve as AM depth increases, assuming that the internal representation of AM improves with increasing AM depth. In fact, consistent with this notion, the discrimination thresholds at short durations tend to decrease as standard AM depth increases from 0.1 to 0.3 (see Table I). This effect is even more apparent when the thresholds are examined in terms of $\Delta m/m$.

The discrimination thresholds in the present study generally decreased with increasing stimulus duration, suggesting that a temporal-integration process was operating. Viemeister (1970) showed that the threshold for the detection of beats improved by roughly the square root of the number of beats. His result was evaluated in terms of a multiple-looks hypothesis that suggests that changes in performance with duration are determined primarily by the number of "looks" at the amplitude fluctuations. Recently, Viemeister and Wakefield (1991) proposed a multiple-looks model to explain temporal integration. It is of interest to consider whether such a model can explain the present results. According to the multiple-looks model, the AM depth discrimination should improve by the square root of the number of looks if individual looks of equal d' are mutually independent, and the information from the different looks is optimally combined (Green and Swets, 1988). The prediction based on the multiple-looks model is evaluated here in experiment 2.

To apply the multiple-looks model to AM depth discrimination, we must consider how to define a look. It seems reasonable to assume that one well-defined peak and valley in a modulation cycle defines a look. However, except for the 10-Hz modulation rate, subjects were not able to do the task at the duration corresponding to one cycle of modulation (see Fig. 2). This is likely due to the fact that a well-defined peak and valley generally did not exist within one complete modulation cycle, because the stimuli were shaped by a cosine-squared function with a 5-ms rise-fall time. In order to achieve one "clean look," given the rise-fall time, it may be necessary to present more than one modulation cycle. Because only an integer number of cycles was presented in the present study, this suggests that the number of looks may be defined as $n - 1$, where n represents the number of modulation cycles. Thus, two cycles of modulation may have been necessary to achieve one optimum look. Consistent with this definition are pilot data from one subject (S1) showing that the discrimination threshold was almost the same for durations corresponding to 1.25, 1.5, and 2.0 cycles of modulation at a 20-Hz modulation rate (this subject could not do the task for a duration corresponding to 1.0 cycle).

II. EXPERIMENT 2: PSYCHOMETRIC FUNCTIONS FOR AM DEPTH DISCRIMINATION

The second experiment examined the extent to which the multiple-looks model can explain the effect of stimulus

duration on AM depth discrimination. In order to do that, psychometric functions were measured at different stimulus durations, and comparisons between obtained d' and predicted d' based on the multiple-looks model were made.

A. Method

1. Subjects

Two subjects (S2 and S3) from experiment 1 participated.

2. Stimulus

Stimulus generation was identical to the previous experiment. A standard depth of 0.18 was used. Two modulation rates were used: 10 and 20 Hz. For each modulation rate, psychometric functions were measured at different stimulus durations. For the 10-Hz modulation rate, the stimulus duration was 200, 400, or 800 ms. For the 20-Hz modulation rate, the stimulus duration was 100, 200, 400 or 800 ms.

3. Procedure

To obtain a psychometric function for each condition, several signal modulation depths were chosen that would cover the range of percent correct values between 50% and 100%. As in the first experiment, a two-interval forced-choice paradigm with feedback was employed. The signal modulation depth was fixed within a block of trials. For each condition, the resulting percent correct values for five 60-trial blocks were averaged and taken as the percent correct for a particular modulation depth.

4. Fitting the psychometric function

Each psychometric function was fitted with a cumulative Gaussian probability function (for details, see Dai, 1995). The detectability (d') is defined as

$$d' = \left(\frac{x}{\alpha} \right)^{\beta}, \quad (2)$$

where α is the signal strength at threshold ($d' = 1$ when $x = \alpha$), and x is defined as $\Delta m/m$. β is the slope of the psychometric function because $\log(d') = \beta \log(x/\alpha)$. The parameters, α and β , were estimated by using a Simplex procedure (the FMINS function of MATLAB software).

B. Results

Figure 4 shows the measured psychometric functions for the different stimulus durations for the two subjects. Percent correct is plotted as a function of $20 \log(\Delta m/m)$. The psychometric functions in the left column are for a modulation rate of 10 Hz, and those in the right column are for a modulation rate of 20 Hz. The upper panels are for S2 and the bottom panels are for S3. In each panel, different symbols represent different stimulus durations. The lines are the fitted functions described above.

In general, within a panel, the psychometric functions overlap (S2) or nearly overlap (S3) at durations yielding four or more cycles of modulation, as expected based on the results from experiment 1 (see Table I). The psychometric

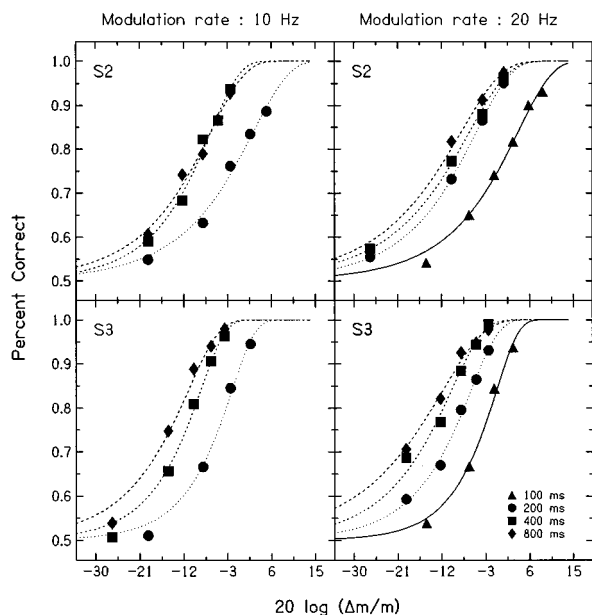


FIG. 4. Psychometric functions of two subjects (S2 and S3) measured at different durations for the modulation rates of 10 Hz (left column) and 20 Hz (right column). The standard depth was 0.18. The durations were 100 (triangles), 200 (circles), 400 (squares), or 800 ms (diamonds).

functions for the durations corresponding to less than four cycles are shifted to the right, to larger values of Δm . Table II shows comparisons between thresholds obtained with the adaptive procedure (from experiment 1) and those obtained with the fixed-level procedure (from experiment 2); in both cases, thresholds correspond to 79.4% correct. These values are not significantly different for both the 10-Hz ($F_{(1,6)} = 0.262$, $p > 0.1$) and 20-Hz rate ($F_{(1,8)} = 0.19$, $p > 0.1$).

All estimated parameters for the fits to the functions are shown in Table III. There is no systematic change in the slope (β) with stimulus duration for S2, but for S3 the slope is steeper at shorter durations.

C. Discussion

The primary purpose of experiment 2 was to examine the extent to which the multiple-looks model might explain the effect of stimulus duration on AM depth discrimination. This can be done in the context of Fig. 5, which shows obtained (unfilled symbols) and predicted (filled symbols) d' as function of stimulus duration. The predicted function was derived based on several assumptions: (a) one well-

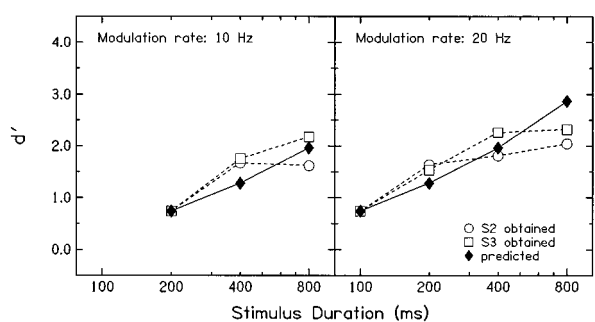


FIG. 5. The obtained (unfilled symbols) and predicted (filled symbols) values of d' as a function of stimulus duration at the modulation rates of 10 Hz (left panel) and 20 Hz (right panel).

defined peak-valley pair provides one “look” (note that, given the rise–fall used here, two cycles of modulation provide one look, and that in general, the number of looks is given by the number of cycles minus 1), (b) individual looks are mutually independent, (c) d' is the same for each look, and (d) the looks are optimally combined. According to these assumptions, the predicted d' should improve by the square root of the number of looks (Green and Swets, 1988). For the present analysis, a value of $\Delta m/m$ corresponding to $d' = 0.74$ (70% correct)² was obtained from the fitted function for the condition providing one look (200 ms for 10 Hz and 100 ms for 20 Hz). That value was used to obtain d' values from the fitted functions for the other conditions.

As can be seen in Fig. 5, the multiple-looks model generally provides a reasonable prediction of AM depth discrimination as a function of stimulus duration. The obtained d' , however, was underpredicted for both subjects as the looks increased from one to three (from 200 to 400 ms for the 10-Hz rate and from 100 to 200 ms for the 20-Hz rate). One possible explanation for this is that the d' for the first look is lower than that for subsequent looks, leading to a greater improvement with increasing looks than predicted based on the d' for the first look. This first look may not be optimal for this task because of adaptation or onset uncertainty (Viemeister, 1979; Viemeister and Wakefield, 1991). Short-term adaptation might reduce sensitivity to the first look relative to later looks if the initial neural response to the onset of the stimulus is large relative to the change in neural response produced by the sinusoidal modulation (see Viemeister, 1979). Onset uncertainty might also reduce the sensitivity to the first look if the subjects do not know precisely when the signal starts (see Viemeister and Wakefield, 1991).

The change in performance with increasing the number

TABLE II. A comparison between the threshold (Δm) from the fixed-level procedure (experiment 2) and that from the adaptive procedure (experiment 1).

| | 10 Hz | | | | 20 Hz | | | |
|--------|-------|----------|-------|----------|-------|----------|-------|----------|
| | S2 | | S3 | | S2 | | S3 | |
| | Fixed | Adaptive | Fixed | Adaptive | Fixed | Adaptive | Fixed | Adaptive |
| 100 ms | | | | | 0.219 | 0.248 | 0.142 | 0.148 |
| 200 ms | 0.185 | 0.269 | 0.121 | 0.094 | 0.082 | 0.057 | 0.075 | 0.082 |
| 400 ms | 0.071 | 0.105 | 0.057 | 0.037 | 0.068 | 0.086 | 0.046 | 0.062 |
| 800 ms | 0.068 | 0.108 | 0.041 | 0.037 | 0.055 | 0.057 | 0.035 | 0.041 |

TABLE III. Estimated values of α and β from the fits to the psychometric functions for two subjects at different stimulus duration. The standard depth was 0.18.

| | | α | | | β | | |
|-------|--------|----------|-------|-------|---------|-------|-------|
| | | S2 | S3 | Avg. | S2 | S3 | Avg. |
| 10 Hz | 800 ms | 0.315 | 0.192 | 0.254 | 0.802 | 0.891 | 0.847 |
| | 400 ms | 0.337 | 0.277 | 0.307 | 0.953 | 1.113 | 1.033 |
| | 200 ms | 0.849 | 0.591 | 0.720 | 0.775 | 1.183 | 0.979 |
| 20 Hz | 800 ms | 0.246 | 0.156 | 0.201 | 0.683 | 0.667 | 0.675 |
| | 400 ms | 0.306 | 0.213 | 0.260 | 0.719 | 0.860 | 0.790 |
| | 200 ms | 0.377 | 0.358 | 0.368 | 0.800 | 0.994 | 0.897 |
| | 100 ms | 1.011 | 0.700 | 0.856 | 0.809 | 1.248 | 1.029 |

of looks beyond four (from 400 to 800 ms for the 10-Hz rate and from 200 to 800 ms for the 20-Hz rate) was overpredicted. Over that range of duration, the obtained d' tended to stay constant or increase slightly with increasing number of looks, whereas the predicted d' increased steadily. This difference between predicted and obtained performance may reflect memory limitations. For longer stimulus durations, the total detectability may be reduced because the detectability of earlier looks has been degraded by memory limitations (Viemeister and Wakefield, 1991). Taken together, it is possible that the detectability of each look may not be equal (the first look might be degraded by adaptation or onset uncertainty, whereas the later looks may be degraded by memory limitations). With appropriate weighting of each look (Viemeister and Wakefield, 1991), however, the multiple-looks model can almost certainly produce more accurate predictions of the integration data for AM depth discrimination. We did not, however, assign an individual weight to each look in order to improve the predictions: because we did not measure the weight for each look directly, this would have been nothing more than an exercise in curve fitting.

III. SUMMARY AND CONCLUSIONS

Experiment 1 measured the discrimination of AM depth as a function of the stimulus duration for different modulation rates. Experiment 2 measured the psychometric functions at different stimulus durations to examine the extent to which the multiple-looks model can explain the effect of stimulus duration on AM depth discrimination. The following conclusions may be drawn.

(1) The threshold for AM depth discrimination decreased by more than a factor of two as stimulus duration doubled from the shortest duration used up to a critical duration, beyond which the threshold decreased only slightly or remained constant. The critical duration corresponds to four cycles of modulation. These results are similar to those for AM rate discrimination (Lee, 1994).

(2) The effect of stimulus duration on AM discrimination was found to be similar to that on AM detection observed previously by others, except at durations shorter than the critical duration. The difference at subcritical durations may be related to the difference in the inherent difficulty of the two tasks.

(3) The multiple-looks model provides a reasonable pre-

diction for AM depth discrimination. A better description, however, could be achieved by assuming different weights for the individual looks.

ACKNOWLEDGMENTS

The authors thank Huanping Dai for his help with the fitting procedure used to fit the psychometric functions, and Joseph W. Hall and an anonymous reviewer for their valuable comments on an earlier version of this manuscript. This research was supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) (DC01376).

¹Intensity compensation was not used here because the modulation depths were very small. Additional measurements with overall level randomized (20-dB range) between intervals confirmed that an overall level cue was not used in the present study: Threshold for the random-level condition was almost the same as that for the fixed-level condition.

²The d' value of 0.74 was chosen arbitrarily. However, the relationship between the predicted and the obtained functions remained essentially the same for initial d' values of 0.545, 0.74, and 1.0.

- ANSI (1989). ANSI S3.6-1989, "Specifications for audiometers" (American National Standard Institute, New York).
- Dai, H. (1995). "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods," *J. Acoust. Soc. Am.* **98**, 3135–3139.
- Grantham, D. W., and Bacon, S. P. (1988). "Detection of increments and decrements in modulation depth of SAM noise," *J. Acoust. Soc. Am. Suppl.* **1** **84**, S140.
- Green, D., and Swets, J. A. (1988). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos).
- Lee, J. (1994). "Amplitude modulation rate discrimination with sinusoidal carriers," *J. Acoust. Soc. Am.* **96**, 2140–2147.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Moore, B. C. J., Glasberg, B. R., Gaunt, T., and Child, T. (1991). "Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals," *Q. J. Exp. Psychol.* **43A**(3), 327–347.
- Ozimek, E., and Sek, A. (1988). "AM difference limens for noise band," *Acustica* **66**, 153–160.
- Sheft, S., and Yost, W. (1990). "Temporal integration in amplitude modulation detection," *J. Acoust. Soc. Am.* **88**, 796–805.
- Viemeister, N. F. (1970). "Auditory discrimination of intensity, internal noise, and temporal processing," Ph.D. dissertation, Indiana University, Bloomington, IN.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.
- Wakefield, G. H., and Viemeister, N. F. (1990). "Discrimination of modulation depth of SAM noise," *J. Acoust. Soc. Am.* **88**, 1367–1373.

Detection of silent intervals between noises activating different perceptual channels: Some properties of “central” auditory gap detection

D. P. Phillips,^{a)} T. L. Taylor,^{b)} S. E. Hall,^{b)} M. M. Carr,^{c)} and J. E. Mossop^{b)}
Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

(Received 7 March 1996; accepted for publication 30 January 1997)

This article describes four experiments on gap detection by normal listeners, with the general goal being to examine the consequences of using noises in different perceptual channels to delimit a silent temporal gap to be detected. In experiment 1, subjects were presented with pairs of narrow-band noise sequences. The leading element in each pair had a center frequency of 2 kHz and the trailing element's center frequency was parametrically varied. Gap detection thresholds became increasingly poor, sometimes by up to an order of magnitude, as the spectral disparity was increased between the noise bursts that marked the gap. These data suggested that gap-detection performance is impoverished when the underlying perceptual timing operation requires a comparison of activity in different perceptual channels rather than a discontinuity detection within a given channel. In experiment 2, we assessed the effect of leading-element duration in within-channel and between-channel gap detection tasks. Gap detection thresholds rose when the duration of the leading element was less than about 30 ms, but only in the between-channel case. In experiment 3, the gap-detection stimulus was redesigned so that we could probe the perceptual mechanisms that might be involved in stop consonant discrimination. The leading element was a wideband noise burst, and the trailing element was a 300-ms bandpassed noise centered on 1.0 kHz. The independent variable was the duration of the leading element, and the dependent variable was the smallest detectable gap between the elements. When the leading element was short in duration (5–10 ms), gap thresholds were close to 30 ms, which is close to the voice onset time that parses some voiced from unvoiced stop consonants. In experiment 4, the generality of the leading-element duration effect in between-channel gap detection was examined. Spectrally identical noises defining the leading and trailing edges of the gap were presented to the same or to different ears. There was a leading-element duration effect only for the between channel case. The mean gap threshold was again close to 30 ms for short leading-element durations. Taken together, the data suggest that gap detection requiring a temporal correlation of activity in different perceptual channels is a fundamentally different task to the discontinuity detection used to execute gap detection performance in the traditional, within-channel paradigm. © 1997 Acoustical Society of America. [S0001-4966(97)02406-5]

PACS numbers: 43.66.Mk, 43.66.Rq [JWH]

INTRODUCTION

One of the most common tasks used for the behavioral measurement of auditory temporal resolution is gap detection. Usually, the listener is presented successively with two long (usually hundreds of ms) stimuli. The “signal” contains a brief (a few ms) silent period at its temporal midpoint, and the “standard” has either no such silent period (“gap”) or else one so short as to be inaudible. The task of the subject is to indicate (e.g., by button press) which one of the two stimuli contained the longer discontinuity or gap (e.g., Moore, 1985). A number of factors influencing gap detection thresholds have been identified. First, studies with bandpassed noise reveal that gap detection thresholds depend more on the bandwidth of the stimulus than its center fre-

quency (e.g., Eddins *et al.*, 1992). This perhaps reflects the greater information transmitted to the central nervous system (after Grose, 1991; Hall *et al.*, 1996). There is agreement that under optimal conditions, i.e., using wideband or high-frequency signals, minimal detectable gaps are in the order of a few milliseconds (Fitzgibbons, 1983; Fitzgibbons and Wightman, 1982; Florentine and Buus, 1984; Plomp, 1964; Shailer and Moore, 1983). Second, there is some evidence that the gap detection performance supported by the apical regions of the cochlea is relatively poor. This is likely because of the greater stimulus uncertainty (i.e., inherent fluctuations in the low-frequency stimulus envelope that might be confused with an intended gap in the stimulus), longer integration times, or slower decision processes within the low-frequency central, perceptual channels, and for extremely low frequencies (less than about 200 Hz), perhaps because the narrow filters of the low-frequency cochlea have longer response times (“ringing”: Moore, 1985; Moore and Glasberg, 1988; Moore *et al.*, 1993).

Note that in the usual gap-detection paradigm, the “tem-

^{a)}Departments of Psychology and Otolaryngology, Electronic mail: ears@is.dal.ca

^{b)}Department of Psychology.

^{c)}Department of Otolaryngology.

poral” task is actually a *discontinuity detection within a perceptual channel*. Information about the stimulus perturbation (offset of the sound that defines the leading edge of the gap and the onset of the sound defining the trailing edge of the gap) can presumably be carried by any or all of the afferent nerve fibers, and their central projections, innervating the cochlea at the locus or loci representing the stimulus content. (In this sense, the term “channel” is not to be confused with the more narrow concept of a perceptual “channel” restricted to a single critical band, after Scharf, 1970.) It is the “within-channel” feature of the processing which renders a neural correlate of gap detection visible in recordings from single cochlear nerve cells (e.g., Zhang *et al.*, 1990).

Many behaviorally important temporal discriminations, however, are not strictly of this kind. In many instances, the gap to be detected is delimited by spectrally different (or nonoverlapping) markers. This requires a *relative timing* operation to be performed on activity *between different perceptual channels*. In the case of discriminating the voice onset time in stop consonants, for example, the task is to judge the relative timing of the (predominantly high frequency) consonantal burst and the subsequent (predominantly low frequency) vowel. Conceptualized within a gap-detection paradigm (e.g., Nelson *et al.*, 1995), the task of the listener in such instances superficially remains the same (“which stimulus combination contains the gap?”), but the mechanisms that mediate the perceptual responses in the two paradigms are likely different. The cochlear nerve array as a whole may contain information about the relative timing of the stimulus elements (e.g., Carney and Geisler, 1986), but it does not contain any machinery capable of executing the relative timing operation, since there are no lateral neural connections between cochlear output fibers. Only if there is significant spectral overlap between the stimulus defining the gap can a single neural channel carry information about the timing of the silent period. To some extent, this situation exists in speech signals (e.g., Sinex and McDonald, 1988; Sinex and Narayan, 1994). In the extreme case of the stimulus elements having no spectral overlap, however, the relative timing operation must presumably be performed centrally.

In this regard, there is behavioral evidence that the introduction of a spectral disparity between relatively low-frequency stimuli defining the leading and trailing edges of the gap in the gap-detection stimulus results in significantly impoverished gap-detection performance (Fitzgibbons *et al.*, 1974; Formby and Forrest, 1991; Formby *et al.*, 1993, 1996; Neff *et al.*, 1982). These findings are consistent with the view that there may be something fundamentally different about the perceptual processes involved in the within-channel and between-channel gap-detection tasks. The purpose of our experiment 1 was to map the extent of this effect systematically, and by so doing, to shed light on the nature of the processes involved in the between-channel, relative timing operation.

The purpose of experiment 2 was to evaluate the consequences of shortening the duration of the leading element in gap-detection stimuli of both the within- and between-channel kinds. Neurophysiological studies of the cortical au-

ditory system have revealed that the coding of gaps is poorer for gaps occurring early (5 ms) rather than late (500 ms) in a noise stimulus (Eggermont, 1995). This suggests that behavioral detection of gaps might follow a similar rule, particularly because behavioral gap detection performance depends on the integrity of the cortical auditory system (Buchtel and Stewart, 1989; Ison *et al.*, 1991; Kelly *et al.*, 1996).

In experiment 3, we studied gap detection performance using stimulus conditions which had the gross properties of a spoken stop consonant, i.e., a broadband leading element, and a predominantly low-frequency trailing one (see also Nelson *et al.*, 1995). We were thus able to use the general gap-detection paradigm as a tool to probe the processes involved in stop consonant discrimination.

In experiment 4, we examined the generality of gap-detection phenomena by studying the effect of variations in leading-element duration in gap-detection stimuli which were spectrally constant (narrow-band noise centered at 4000 Hz), but in which the leading and trailing noise elements were presented to the same or to different ears (see also Penner, 1977).

I. EXPERIMENT 1

The purpose of experiment 1 was to examine the effect on gap-detection thresholds of systematically varying the direction and size of the frequency disparity between the stimuli defining the leading and trailing edges of the gap in a conventional gap-detection paradigm. In order that disparities in excess of two octaves could be used, and that the center frequencies used remained within both the standard audiometric range and the dynamic range of our stimulating system, the leading noise burst was always centered on 2.0 kHz.

A. Methods

1. Subjects

The authors (four female, one male) participated as listeners. All of the subjects had auditory thresholds within laboratory norms, and three of the listeners, who participated in a separate study, were shown to have normal hearing sensitivity to 8000 Hz by a clinically certified audiologist. Three subjects (SH, JM, DP) had significant prior experience in auditory perception studies.

2. Apparatus, stimuli, and procedure

Stimulus presentation and data acquisition were controlled by a custom programmed Macintosh IIx computer. All stimuli were 0.25-oct, bandpassed noise (48 dB/oct), generated by filtering the output of a white-noise source (Coulbourn Instruments). The noise stimuli were shaped with 0.5-ms linear rise-fall times, including those defining the gaps. The shaped stimuli were passed through a passive attenuator, a power amplifier, and directed to Koss SST/10 digital stereo headphones. Detection thresholds were obtained for all stimulus elements used. All stimulus presentation was monaural, at 40 dB SL (sensation level: dB *re*: the subject’s threshold for that stimulus), and to the listener’s

preferred ear. All testing was done individually, with the subject seated in a sound-attenuating room (Eckel).

Figure 1(a) shows schematically the design of the stimuli used in experiment 1. Each stimulus trial consisted of two sequences of noise bursts. The independent variable was the center frequency of the noise bursts that defined the trailing edges of the gaps. The dependent variable was the minimum detectable gap. The intersequence interval was 300 ms. Each sequence had a total length of 600 ms. One of the sequences (hereafter the “standard” sequence) began with a 0.25 oct-wide noise burst centered on 2.0 kHz. The second element was also a 0.25 oct-wide noise burst whose center frequency was the independent variable (ranging from 0.3 to 8.5 kHz). The two elements comprising the standard were equally long and separated by an inaudible gap (1.0 ms, including rise–fall times). The second (“signal”) sequence had the same overall duration, and the same order and composition of elements. It differed from the first sequence only in that the gap separating the two elements was determined by the adaptive tracking procedure.

Even in within-channel stimulus conditions, both the standard and the signal contained gaps (1.0 ms, in the case of the standard). This made it less likely that listeners could use the presence of gating transients *per se* as a cue to the identity of the signal. Technically, the presence of a silent period in the standard means that the listeners might be construed as performing a gap discrimination (cf. detection) task, but we emphasize that no participant in the study reported detectability of the gap in the standard, in either the within- or between-channel stimulus conditions. In principle, it is possible that listeners were using overall stimulus energy as a cue, but listener reports of their strategies never included mention of perceived loudness, and typically did include descriptions of the “smoothness” or “abruptness” with which the leading noise burst blended into the trailing one. See also experiment 3 (below) for a demonstration that the listeners were probably not using overall stimulus energy to identify the signal.

In a given block of trials, the order of standard and signal was randomized by the computer. The task of the subject was to indicate by button-press whether the first or second stimulus contained the (longer) gap. The intertrial interval was usually about 1000 ms, including a 500-ms warning light. The order of elements in a stimulus was held constant to reduce stimulus uncertainty (after Green and Forrest, 1989). In pilot studies using the same paradigm, we found no effect of element order when subjects were studied with blocks of trials in which the order of elements within the stimuli was counterbalanced (see also Williams and Perrott, 1972).

Each value of the independent variable was tested in a separate block of trials using a two-down/one-up method (after Levitt, 1971). Briefly, the starting duration of the signal gap was large (around 20 ms for conditions of a small frequency disparity, and as high as 220 ms for conditions in which there was a large disparity), and after every two successive correct responses, it was decreased by a factor of 1.2. After an incorrect response, the size of the gap was increased by the same factor. This two-down/one-up procedure contin-

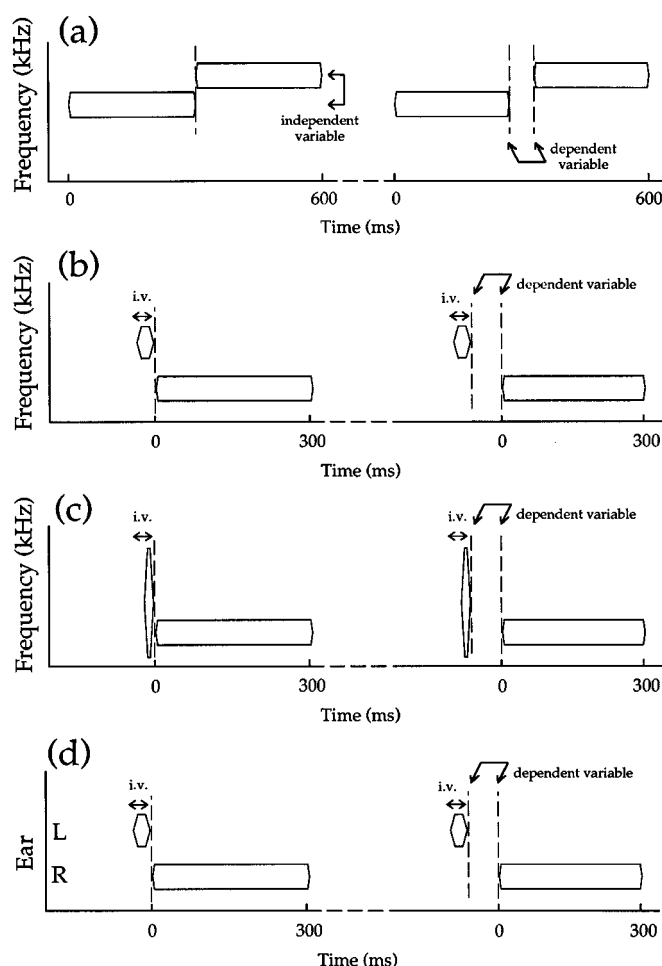


FIG. 1. (a) Schematic diagram depicting stimulus design in experiment 1. Each trial consisted in the presentation of two sequences of noise bursts, each of whose durations was 600 ms. Each sequence began with a 0.25-oct-wide noise burst with a center frequency of 2.0 kHz. The center frequency of the trailing noise burst was the independent variable. The standard contained an undetectable silent period (gap) of 1 ms. The signal contained a longer gap, and the duration of this gap was the dependent variable. The order of standard and signal was randomized across trials. The task of the listener was to determine which sequence was the signal. (b) Schematic diagram depicting stimulus design in experiment 2. Each trial consisted in the presentation of two sequences of noise bursts. In the within-channel conditions, the noises were narrow-band and centered at 4.0 or 1.0 kHz (in separate blocks of trials); the duration of the leading element was the independent variable. In the between-channel conditions, the leading element was always the high-frequency noise, and the trailing noise was always the low-frequency one. In the standard, the two elements were separated by an undetectable gap of 1.0 ms. The duration of the silent period in the signal was the dependent variable. The order of standard and signal was randomized across trials. The task of the listener was to indicate which sequence was the signal. (c) The general design of experiment 3 was the same as that of experiment 2, except that the leading element was always a wideband noise, and the trailing element was always a narrow-band noise centered on 1.0 kHz. There were no within-channel conditions in this experiment. (d) The general design of experiment 4 was the same as that of experiment 2, except that all stimulus elements were 4.0 kHz narrow-band noises. The leading element in both the standard and the signal was always presented to the left ear. In separate blocks of trials, the left leading element was paired with the trailing one which was delivered to the left, the right, or to both ears simultaneously.

ued until there had been 14 reversals in the direction of change of the test gap size. The gap threshold for that block of trials was taken as the mean gap duration across the last ten trials that produced reversals in the adaptive step. Listen-

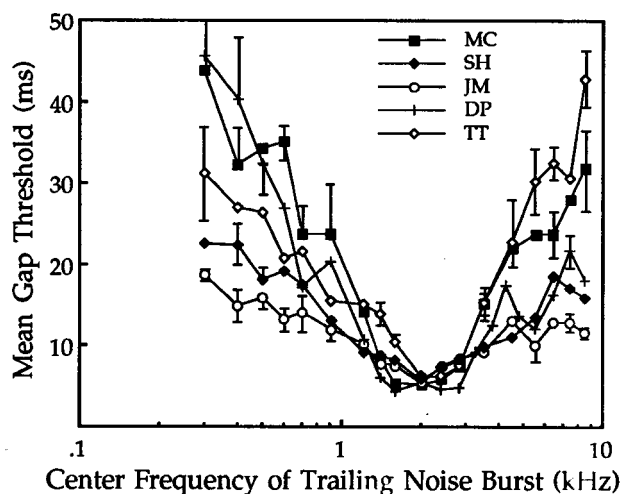


FIG. 2. Data obtained in experiment 1, shown separately for each listener. Mean gap thresholds are plotted as a function of the center frequency of the trailing noise burst used in the test and reference sequences. Note that while all of the functions were broadly V-shaped, three listeners showed a tendency for gap thresholds to plateau for high values of the independent variable. Sample standard errors are shown.

ers typically took 40–60 trials to reach this criterion. Over a period of three to four weeks, every stimulus condition was tested repeatedly until gap thresholds were relatively stable. Subjects took between six (subject MC) and ten blocks (subject DP) at each stimulus condition to reach consistently best performance. In practice, learning curves for the within-channel conditions were very short; there was little variance in gap thresholds within subjects across blocks. For the between-channel conditions, learning curves were longer; subjects often took twice as many blocks to reach plateau performance.

B. Results and discussion

Data collected were mean durations of minimal detectable gaps over the last four blocks of trials for each value of the independent variable. These data are plotted separately for each of the five subjects in Fig. 2, which shows mean gap threshold as a function of the center frequency of the noise burst defining the trailing edge of the gap. Sample standard error bars are shown; those for stimulus conditions with trailing noise bursts with center frequencies at or near 2.0 kHz are too small to be visible.

There were considerable individual differences in the fine structure of these functions. The sample standard errors seen in Fig. 2 are relatively small by comparison with the individual differences in gap-detection thresholds for given values of the independent variable. The reasons for the size of these individual differences are not clear, but may reflect the consequences of any individual differences in loudness growth across the stimulus frequencies used. Notwithstanding this point, the five curves have a number of gross features in common. First, gap thresholds were smallest for stimulus conditions in which the center frequency of the trailing noise burst was at, or very close to, that of the noise burst defining the leading edge of the gap. Minimal detect-

able gaps for the within-channel (2.0 kHz/2.0 kHz case) were 5.3, 5.3, 5.4, 5.8, and 6.3 ms.

Gap-detection thresholds depend on the number of blocks tested (and thus amount of practice), stimulus bandwidth (0.25 oct in this study, but constant in Hz in other studies), and conservatism of the threshold tracking strategy (two-down/one-up in this study; three-down/one-up in Eddins *et al.*, 1992). The present values for the within-channel case at 2.0 kHz are similar to (Moore and Glasberg, 1988; Neff *et al.*, 1982) or smaller than those reported by other authors working with stimuli in the same frequency range (Formby and Muir, 1988; Formby and Forrest, 1991; Formby *et al.*, 1993; Grose *et al.*, 1989; Eddins *et al.*, 1992). It is possible that some feature of our stimulus design permitted listeners to use unintended spectral or other cues to identify the signal, and that this is responsible for the relatively low gap thresholds seen in this study. It is also possible, however, that the extended practice of the present listeners, the more liberal threshold tracking strategy, and the broader stimulus bandwidths in this experiment all favored low, and uncontaminated, gap thresholds.

Second, for each subject, gap threshold was roughly a V-shaped function of the independent variable. The present subjects, who performed as well as published norms for the within-channel case, were between three and ten times worse when there was a two octave disparity between the center frequencies of the noise bursts defining the gaps. These data are broadly comparable to those of Neff *et al.* (1982) who used tonal sequences centered around 1600 Hz. The reasons for the strength of the individual differences are unclear. It remains the case, however, that for each subject, gap thresholds for gaps delimited by noise markers with medium and wide frequency disparities were virtually (or totally) non-overlapping with those seen in the within-channel case.

Third, for all subjects, on the low-frequency side of the functions, gap thresholds systematically increased with increasing disparity between the center frequencies of the noise bursts defining the gap. The data are less clear for the high-frequency slopes of the functions shown in Fig. 2. Two of the subjects (MC, TT) showed gap thresholds that more or less continuously increased with increasing frequency disparity. The remaining subjects (SH, JM, DP), those with more extensive psychophysical experience, showed some evidence of a plateau for trailing noise-burst center frequencies of 4.0 kHz and above. For these three listeners, the poorest gap thresholds on the high-frequency side of their functions were as much as 1.5 to 2.0 times better than the poorest gap thresholds seen on the low-frequency side for frequency disparities of comparable magnitude on the two sides. In this regard, while subject MC did not show clear evidence of a plateau on the high-frequency side of her function, her gap thresholds were nonetheless slightly better on that side than on the low-frequency one. The asymmetric pattern of these gap detection results for the low- and high-frequency sides of the functions in Fig. 2 is consistent with those measured and modeled for the detection of silent gaps in sinusoidal markers (Formby *et al.*, 1996; Forrest and Formby, 1996; Heinz *et al.*, 1996).

II. EXPERIMENT 2

The data from experiment 1 indicated that between-channel gap-detection thresholds were poorer than those for within-channel gap detection. In the gaps of interest in speech sounds, e.g., voice onset times, the leading element not only has a different spectrum to the trailing one, but it is also shorter in duration, in the range of a few milliseconds (e.g., Stevens and Blumstein, 1978). Experiment 2 was therefore designed to explore the effects of manipulating the duration of the leading element in the gap-detection stimulus. The question is whether there is an effect of leading-element duration on gap thresholds, and whether such an effect is common to both within- and between-channel gap detection. In this regard, there is evidence that behavioral gap-detection performance depends on an intact cortical auditory system (Buchtel and Stewart, 1989; Ison *et al.*, 1991; Kelly *et al.*, 1996). This may be an important point because there is also cortical neurophysiological evidence of poorer coding of gaps occurring early (5 ms) rather than late (500 ms) in spectrally homogeneous noise bursts (Eggermont, 1995). The purpose of experiment 2 was explicitly to examine the effect of leading-element duration on gap-detection performance separately for high-frequency noise bursts, low-frequency noise bursts, and for the between-channel case in which the leading element was high frequency and the trailing element was low frequency.

A. Methods

1. Subjects

Three of the listeners who participated in experiment 1 served as subjects in the second experiment (SH, DP, TT).

2. Stimuli and procedure

The apparatus was the same as in experiment 1. The general design was that the leading noise element duration was the independent variable (and took values of 5, 10, 30, 50, 100, and 300 ms), while the trailing element always had a duration of 300 ms. Each subject was tested under three conditions, viz, with low-frequency bandpassed noise (1.0 kHz center frequency, 0.25 oct bandwidth), with high-frequency noise (0.25-oct, 4.0-kHz center frequency), and using a between-channel design in which the leading element was the 4.0-kHz noise, and the trailing element was the 1.0-kHz noise. In order to mimic the design of the neurophysiological studies (Eggermont, 1995), no attempt was made to match the sensation levels of all the stimulus elements used. Stimulus levels were set independently at 40 dB SL for the 300-ms noise stimuli at 4.0 and 1.0 kHz; the same attenuator settings for each stimulus frequency were used in all stimulus conditions. As a consequence, the 5- and 10-ms duration leading elements were subjectively quieter than the longer elements, but the stimuli remained highly audible for all listeners in all stimulus conditions.

B. Results and discussion

Data entered into the analysis were the means of the last four thresholds obtained in each stimulus condition, after each subject's performance had reached a plateau (which

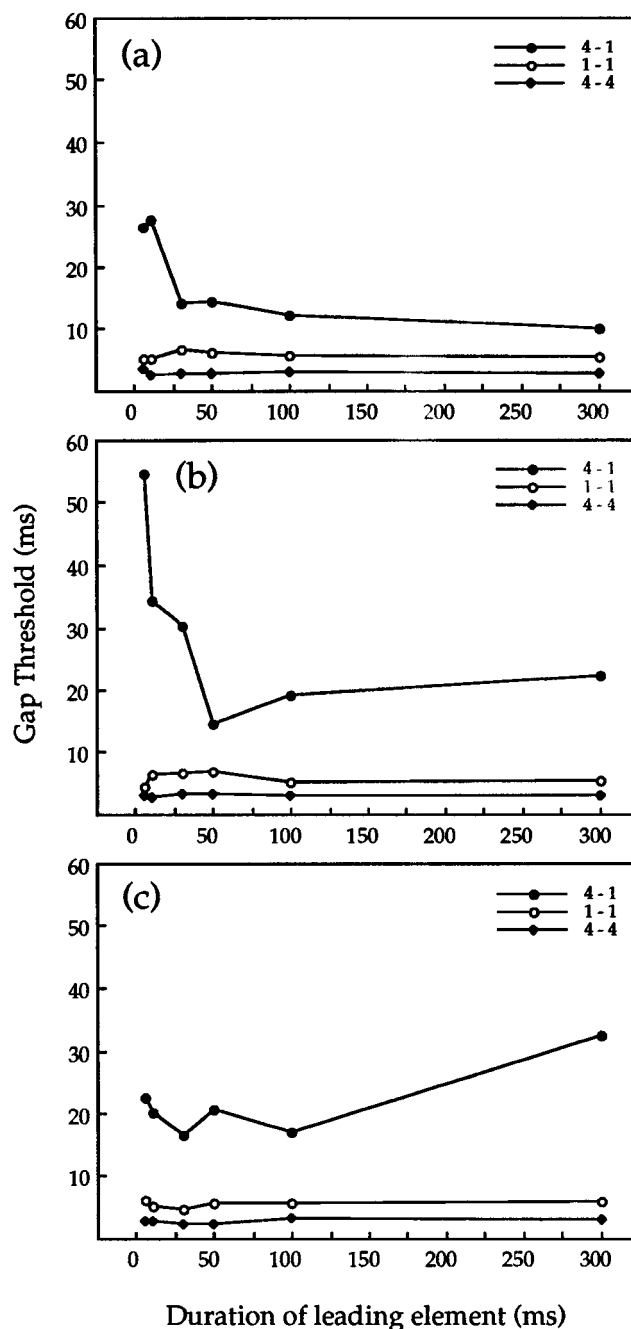


FIG. 3. Data obtained in experiment 2. Each panel presents data for one listener: (a) subject SH, (b) subject DP, (c) subject TT. Mean gap threshold is plotted as a function of the duration of the leading stimulus element; parameter is the spectral condition.

took as few as five threshold determinations in the within-channel cases, and up to 12 threshold determinations for the between-channel cases). Figure 3(a)–(c) shows the data for the three listeners presented separately. The minimal detectable gap duration is plotted as a function of leading-element duration in each panel. The parameter is the spectral condition. Standard errors are not shown, but for most of the leading-element durations, the distributions of individual thresholds for the three spectral conditions were nonoverlapping, or almost so, for each listener.

Gap thresholds for each subject were smaller for the 4.0-kHz condition than for the 1.0-kHz condition, irrespec-

tive of the leading-element duration. The absolute bandwidth of the 4.0-kHz stimulus was greater than that for the 1.0-kHz stimulus. The superior performance for the high-frequency condition is thus consistent with earlier reports of superior gap-detection thresholds for broadband stimuli (Eddins *et al.*, 1992; Green and Forrest, 1989; Grose, 1991; Hall *et al.*, 1996). The absolute values of the gap thresholds were, like those in experiment 1, comparable to, or lower than, those reported previously (Formby and Muir, 1988; Formby and Forrest, 1991; Moore and Glasberg, 1988; Eddins *et al.*, 1992), presumably for the same reasons. The second point to be gleaned from Fig. 3 is that, for each subject, performance was poorer for the between-channel case than for either of the within-channel cases. This observation is consistent with the findings of experiment 1.

The most important data in Fig. 3 pertain to the fact that there was no systematic effect of leading-element duration on gap detection in either of the within-channel cases in any of the listeners, while there was a dramatic lengthening of gap thresholds for subjects SH [Fig. 3(a)] and DP [Fig. 3(b)] when leading-element durations were short. The data for subject TT [Fig. 3(c)] were less orderly, largely because she had a poor gap threshold for the 300-ms condition in the between-channel case.

Penner (1977) described a systematic lengthening of the gap thresholds when the leading elements were increased from 2 to 200 ms, an effect which she argued was due to rate of decay of sensation. In her experiment, however, the trailing element was always 2 ms in duration, so that her experiment is not directly comparable with ours. In the present experiment, the spectrum levels of the leading elements were constant, because we did not covary the stimulus level with its duration to ensure a constant SL for the leading element. In this regard, the *sensation* levels of the leading elements in the present experiment covaried with their durations, at least for the shortest values of the independent variable. This raises the question of whether the leading-element duration effect in the present between-channel condition was in some way due to this factor. Following the general arguments of Plomp (1964) and Penner (1977), such is very unlikely to be the case. This is because a decay-of-sensation hypothesis would predict longer gap thresholds for *longer* leading elements, which is the opposite of what we observed empirically. It is also unclear why a decay-of-sensation hypothesis would predict a leading-element duration effect only in the between-channel case.

III. EXPERIMENT 3

Experiment 1 revealed that between-channel gap thresholds were longer than within-channel thresholds. Experiment 2 revealed that shortening the duration of the leading element in the gap-detection stimulus resulted in lengthening of minimal detectable gaps, but only in the between-channel case. In experiment 3, the stimulus elements used to define the gap were wideband noise (leading element), intended to simulate a consonantal burst, and narrow-band, low-frequency noise (trailing element), intended to simulate a vowel. As in conventional speech materials then, although there was spectral overlap between the leading and trailing elements, the stimu-

lus paradigm was predominantly a between-channel one. The general design, illustrated schematically in Fig. 1(c), was similar to that of the between-channel case in experiment 2.

A. Methods

1. Subjects

Three of the listeners who participated in experiment 1 (SH, DP, TT) served as subjects in experiment 3.

2. Stimuli and procedure

The apparatus was the same as in experiment 1. Each stimulus trial again consisted of two noise sequences: A standard which began with a leading wideband (20 kHz) noise, followed by an inaudible (1.0 ms, including 0.5 ms rise–fall times) gap, and subsequently by a 300-ms, 0.5 oct-wide bandpassed noise centered at 1.0 kHz (filter roll-off was 48 dB/oct). The signal was identical, except that the gap between the elements was long enough to be audible. The independent variable was the duration of the leading wideband element in the two sequences (durations used were 5, 10, 30, 50, 100, and 300 ms). The dependent variable was the minimum detectable gap in the signal.

As in experiment 1, all testing was executed individually. Detection thresholds were obtained separately for all stimulus elements used (six durations of broadband noise and the 300 ms narrow-band noise), and all stimuli in the experiment were presented monaurally to the listener's preferred ear at 40 dB SL. The same two-down/one-up procedure was used to determine the gap threshold for each value of the independent variable. Each subject was tested repeatedly over a period of 3 to 4 weeks until gap thresholds had stabilized for each value of the independent variable. Gap threshold for each subject was taken as the mean of those obtained in the last four blocks of trials.

In this paradigm, the overall duration of the noise sequences used varied with the length of the test gap [see Fig. 1(c)]. As a control procedure, we separately obtained gap thresholds for the wideband-noise/1.0 kHz-noise between-channel case using the design of experiment 1 (in which the overall length of the sequences did not vary with the duration of the gap). Using the experiment 1 design, mean gap threshold was 10.3 ms. Using the experiment 2 design, with a 300-ms leading element, mean gap threshold was 9.8 ms. These were not significantly different by analysis of variance ($F < 1$).

B. Results and discussion

Figure 4(a) shows mean gap thresholds plotted as a function of the leading-element duration; the data are shown separately for the three listeners. Figure 4(b) shows the grand mean thresholds. For each subject, gap thresholds were longest for the shortest leading-element durations, and shortest for the longest leading-element durations. For the shortest values of the independent variable (5 and 10 ms), the mean gap detection threshold was 32.4 ms. For long leading-element durations, mean gap thresholds were close to 10 ms. Thus, in this experiment, we observed an inverse relation between the size of the gap detection threshold and the size

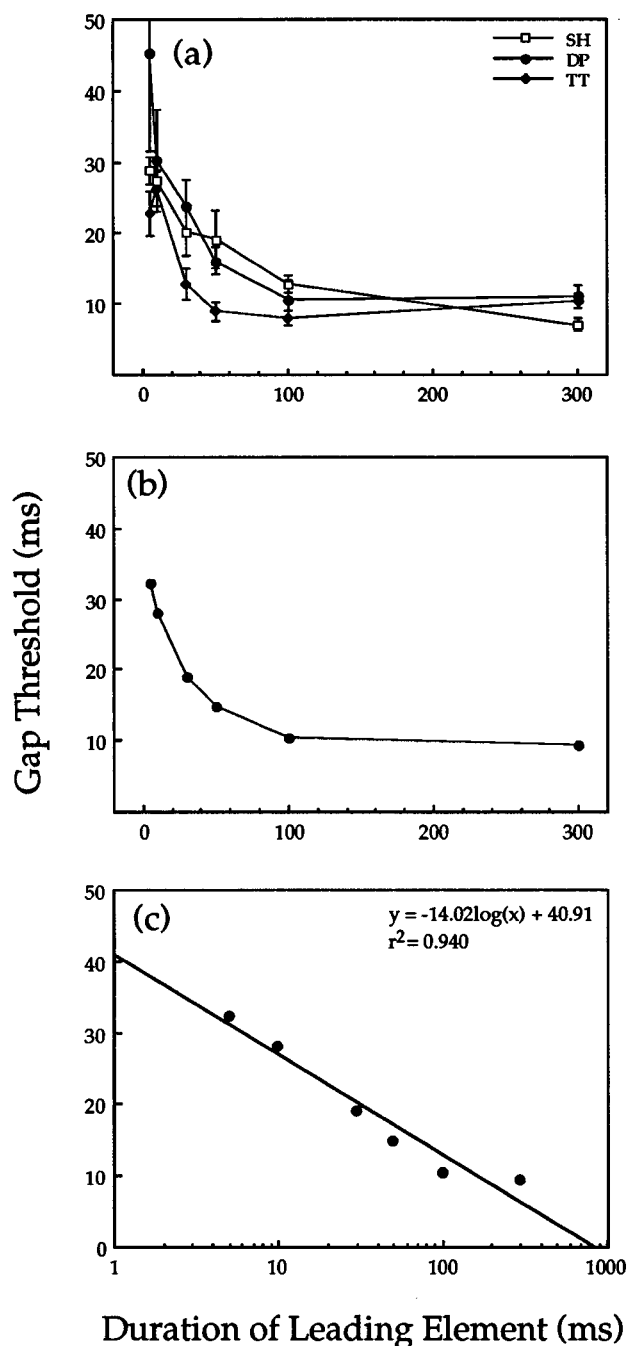


FIG. 4. Data obtained in experiment 3. (a) shows mean gap thresholds plotted as a function of the duration of the leading element, and shown separately for the three listeners. (b) shows the grand mean across the listeners. Crossbars are standard errors. Note that both the mean curve, and the curves for each listener plotted separately were continuous, with gap thresholds declining toward an asymptotic value near 10 ms for long values of the independent variable. (c) shows the same grand mean scores, but displayed on a semilog plot, with a least-squares linear regression line fitted to the data. The linear regression accounted for 94% of the data variance.

of the leading-element duration. These results are consistent with those of the between-channel case in experiment 2.

The long (10 ms) gap thresholds obtained with long leading-element durations are perhaps surprising, because the leading marker was broadband. One explanation for this pattern resides in the spectral differences of the markers. The offset of the silent period was marked by the onset of a

low-frequency noise, and because the bulk of the spectral energy in the leading element was outside this bandpass, the perceptual operation performed to detect the gap was predominantly a between-channel cross correlation (after the fashion of experiment 1, above). This speculation is in keeping with two other studies which revealed increases in gap thresholds for spectrally dissimilar markers designed to resemble speech components (Formby *et al.*, 1993; Nelson *et al.*, 1995). Perhaps more importantly, experiment 3 revealed a major effect of leading-element duration, which experiment 2 previously revealed was a property only of between-channel gap detection. Parenthetically, note that in experiment 3, all stimulus elements were presented at 40-dB sensation level. The fact that a leading-element duration effect was seen here confirms our earlier argument that the leading-element duration effect seen in experiment 2 was not due to the varying sensation levels of the leading elements in experiment 2.

There is, however, a second possibility. In the strictly within-channel gap detection tasks, gap thresholds were about 3–7 ms (experiments 1 and 2). In the strictly between-channel task of experiment 2, gap thresholds ranged from 10 to 33 ms, and were about 23 and 33 ms for DP and TT. In experiment 3, gap thresholds for the longest leading-element durations are about 10 ms. This is much shorter than the gap threshold for the between-channel case in experiment 2. Accordingly, for long leading-element durations in experiment 3, the within-channel comparison may be dominant. The same cannot be said for the shorter leading-element durations. It is thus possible that as leading-element duration in experiment 3 was shortened, the task switched from a within-channel one to a between-channel one.

The stimuli in experiment 3 were designed to resemble, at least grossly, the structure of stop consonants, i.e., a relatively short, broadband leading element (resembling a consonantal burst), followed by a brief period of silence (voice onset time, VOT), and then a relatively long, steady-state, low-frequency element (vowel). For leading elements as short as actual consonantal bursts (a few milliseconds), experiment 3 revealed that the shortest detectable gaps were in the order of 30 ms. What makes this value of special interest is that it is close to the value defining the categorical boundary between some voiced and unvoiced stop consonants (about 25–35 ms, e.g., Eggermont, 1995; Eimas and Corbit, 1973; Kuhl and Miller, 1978; Morse and Snowdon, 1975).

This correspondence raises the intriguing possibility that the categorical boundary has the value it does because it separates inaudible gaps (VOTs) from audible ones, at least for very short leading elements. This line of argument is consistent with Kuhl and Miller's (1978) view that categories in speech sounds exploit "natural psychophysical boundaries." Given the two-down/one-up procedure used in the present experiments, collapsing percent-correct scores across test-gap duration results in a sigmoidal curve resembling a VOT labeling function with a boundary close to 30 ms. Whether redesigning experiment 3 into a discrimination study would result in a discrimination function that peaks for gap durations near 30 ms remains to be seen, as does the issue of whether the minimal detectable gap for the 5-ms

leading-element condition can be altered by selective adaptation procedures (after Eimas and Corbit, 1973).

A qualification of this general argument is that the VOT which parses voiced from voiceless consonants varies with the place of articulation (Kuhl and Miller, 1978). In this regard, however, so too does the spectral composition of the consonantal burst (Stevens and Blumstein, 1978), and so too might the duration of the consonantal burst. These are important points, because experiment 1 revealed an effect of spectral similarity of the noise markers on gap detection thresholds, and experiment 3 revealed that for short leading elements, small changes in leading-element duration also caused large changes in gap-detection thresholds. The fact that VOT boundaries vary with place of articulation may not, therefore, be problematic to Kuhl and Miller's (1978) argument, or to our extension of it.

Finally, experiment 3 provided an opportunity to examine an account of attentional processes mediating gap-detection performance for silent periods defined by spectrally dissimilar markers. Fitzgibbons *et al.*'s (1974) general model suggests that perceptual or attentional resources are initially directed at the frequency channel representing the first element of the sequence. The model suggests that there may be a minimum attentional dwell time on this channel before there can be a time-consuming redirection of perceptual resources to the channel representing the second element. The detection of the gap can presumably occur only if the dwell time is exceeded by the combined durations of the leading element and the gap. It follows that if the only independent variable in the gap-detection paradigm is the duration of the leading element, then for long leading-element durations, the minimal detectable gap will represent only the shift time. As the duration of the leading element is shortened, the minimum detectable gap will increase to the extent that the dwell time exceeds the duration of the first element and the gap. Since the size, direction, and number of shifts remain constant, an estimate of the duration of the dwell time will be represented as the leading-element duration at which the gap-threshold/element-duration function develops a nonzero slope. In a serial model of this kind, the slope of the function seen at short leading-element durations should be close to (negative) unity, and the entire function should resemble the intersection of two straight lines.

Empirically, however, the relationship between gap threshold and leading-element duration did not resemble the intersection of two straight lines, but rather, appears as a single, smooth curve with no obvious discontinuities. This is seen in both the mean data and the curves for the individual listeners. The data thus appear not to be of the form required by a simple, serial-additive model of the kind outlined above.

An alternate interpretation of the data in Fig. 4(b) is that the curve resembles a single exponential recovery function. In Fig. 4(c), the same data points are displayed on a semilog plot, and a least-squares regression line was fitted to them. The fitted straight line accounted for 94% of the data variance, which is consistent with the view that the form of the function reflects a single process.

IV. EXPERIMENT 4

The purpose of experiment 4 was to examine the generality of the phenomena seen in between-channel gap detection. In the present experiment, subjects performed gap detection on a narrow-band noise stimulus in which the leading element was presented to one ear, while the noise defining the offset of the gap was presented to the same, the opposite, or to both ears. In the between-channel paradigm of this study, the relative timing operation required to define the gap had to be executed on perceptual channels dictated not by stimulus spectrum, but by stimulus laterality.

A. Methods

1. Subjects

Three of the listeners who participated in experiment 1 (SH, DP, TT) served as subjects in experiment 4.

2. Apparatus, stimuli, and procedure

The apparatus and general methods were the same as in the preceding experiments. A schematic illustration of the general stimulus design is shown in Fig. 1(d).

There were two independent variables in the experiment. One was the laterality of the trailing noise burst. The leading noise burst was always presented to the left ear. In separate blocks of trials, the trailing noise burst was presented to the left, to the right, or to both ears simultaneously. In what follows, we refer to these as left-left, left-right, and left-both conditions, respectively. The second independent variable was the duration of the leading element, which had values of 5, 10, 30, 50, 100, and 300 ms. The duration of the trailing element was always 300 ms. In this three-by-six within-subjects design, we were able to ascertain the effect of leading-element duration for each pairing of leading and trailing noise lateralities.

Detection thresholds were obtained for 300-ms stimulus durations for each listener, for each ear separately. All subsequent stimulus presentation was with an attenuator setting appropriate for 40 dB SL for the 300-ms stimulus. As a consequence, the sensation levels of the shortest stimulus elements were reduced by 7–10 dB relative to the sensation levels of the longer stimuli. Despite the reduced sensation levels, they were always clearly audible. Note, however, that experiments 2 and 3 (above) had suggested that compensating for such loudness differences was inconsequential to the demonstration of the between-channel gap-detection phenomena. Each listener was tested with each stimulus condition between four and ten times, until their gap thresholds stabilized. Final gap thresholds were the arithmetic means of those obtained in the last four blocks of each stimulus condition.

B. Results and discussion

Mean gap thresholds are shown in Fig. 5, plotted as a function of leading-element duration, and separately for the three laterality conditions. The individual data are shown for three listeners in panels (a)–(c), and the average data for the group are shown in panel (d). For each listener, gap thresholds for the left-left and left-both conditions were low, near

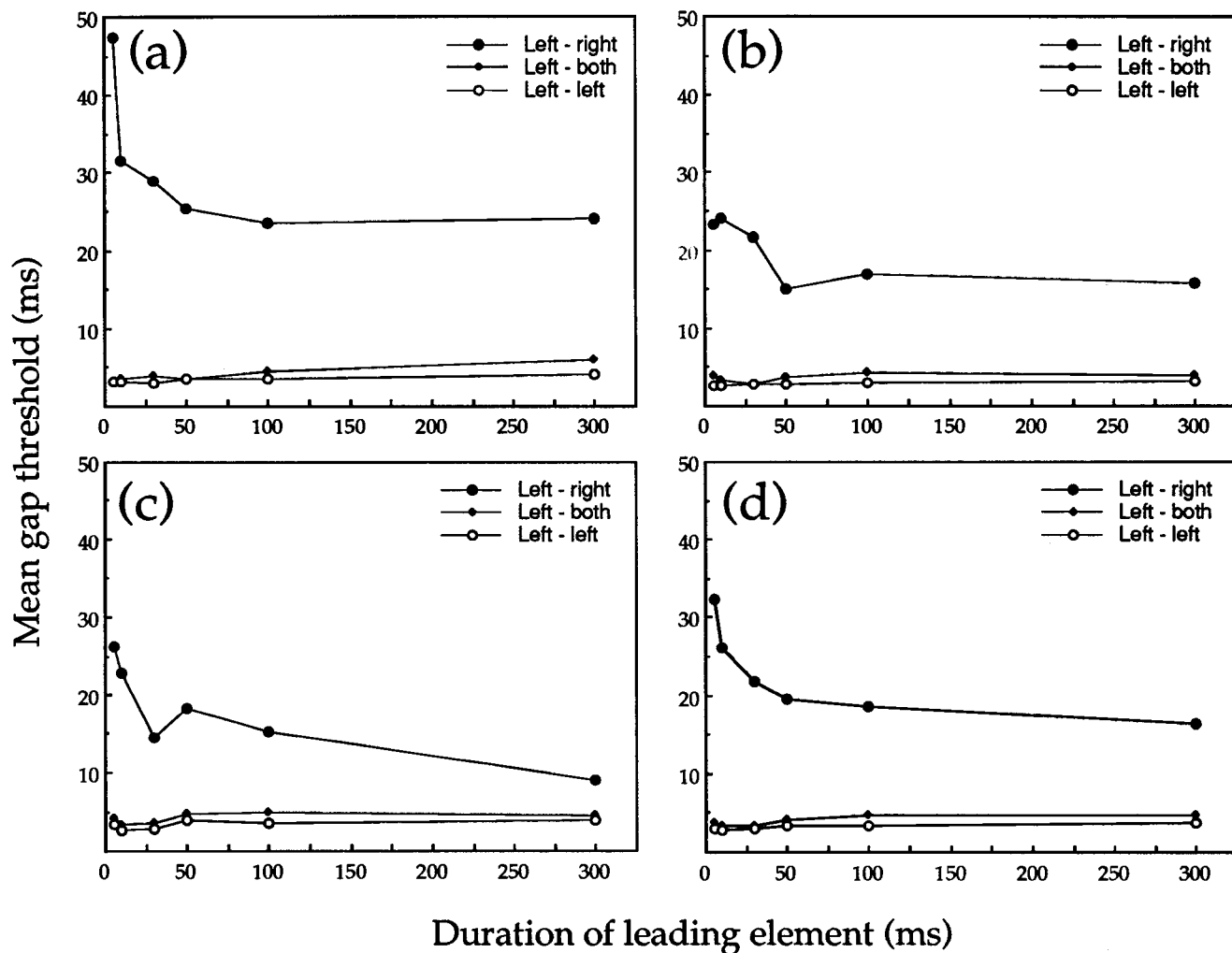


FIG. 5. Data obtained in experiment 4. Panels (a)–(c) show mean gap thresholds, plotted as a function of leading element duration, for each listener (DP, TT, and SH, respectively). Parameter is the laterality configuration of the stimulus. Panel (d) shows the grand mean of the data presented in (a)–(c). Note that gap thresholds were grossly larger in the left–right condition for each listener, and that only in this condition was there an effect of leading-element duration.

4 to 5 ms, and were independent of leading-element duration. There was some tendency for gap thresholds in the left–both condition to be slightly greater than those seen in the left–left conditions, but usually this difference was only a fraction of a millisecond. By contrast, gap thresholds for the left–right conditions were dramatically higher than those for the other two conditions. Gap thresholds, even for long leading-element durations, were between two and five times greater for the left–right condition than for either of the other two laterality conditions. The grand mean curve for the left–right condition is representative of the individual data and reflects an orderly inverse relation between gap threshold and leading-element duration. For the 5-ms leading-element condition, the mean gap threshold in the left–right condition was 32.3 ms and was about 17 ms for the 300-ms leading-element condition.

As in the preceding experiments, it is important to minimize the possibility that listeners used unintended spectral or other cues to identify the gap signal. In the present experiment, both the standard and the signal contained a gap (even though it was not detectable in the standard), and this reduced the likelihood of using the presence of gating tran-

sients *per se* as a cue to which stimulus was the signal. Because the standard and the signal differed in overall duration (by an amount equal to the difference in the durations of their silent periods), it is possible that listeners could use the overall duration of the stimulus as a cue. Gap thresholds in the within-channel conditions were close to 5 ms (Fig. 5). This value represents less than 2% of the total duration of even the shortest standard, and is therefore likely below the difference limen for duration discriminations (about 15% in our preliminary studies). Moreover, listeners described the execution of the within-channel tasks as the detection of a “hiccup” or “stutter” near the onset of the silent gap. In the between-channel conditions, gap thresholds were longer. If listeners had been using overall stimulus duration as a cue, then, following Weber’s Law, gap thresholds should have been longer for stimuli with long leading-element durations. In practice, the reverse was true. Finally, as in experiment 2 above, the fact that the leading elements differed in sensation level cannot account for the leading-element duration effect, (a) because the decay-of-sensation hypothesis (after Plomp, 1964; Penner, 1977) should apply to the within-channel paradigm as much as to the between-channel one, which was not

observed empirically, and (b) because it would predict longer gap thresholds for conditions with longer leading elements, which is the opposite of what was observed.

In the present study, the left–left stimulus conditions were, by definition, within-channel gap detection tasks, and the left–right stimulus conditions were between-channel tasks. The case for the left–both conditions is more ambiguous. Perceptually, the “both” stimulus configuration resulted in a midline percept, and one might therefore have anticipated that the left–both condition was a between-channel one, in which the leading noise burst was of one laterality (left) and the trailing noise burst was of a different laterality (midline) and therefore of a different perceptual channel. In practice, however, gap thresholds for the left–both conditions were very close to those for left–left conditions (Fig. 5), and thus, were not suggestive of between-channel processing. In this regard, what makes the both condition different from the left one is the presentation of stimulus energy to the right ear, but without any compromise of the stimulus at the left ear. This means that in the left–both condition, the presence of the stimulus at the right ear in the trailing noise burst was irrelevant to the execution of a within-channel gap detection on the stimulus at the left ear. Thus, for this task, within-channel (left–left) gap detection, presumably mediated at the auditory periphery, takes precedence over between-channel (left–right) gap detection that must reflect processing at a higher level (or levels) of the auditory system.

The data from experiment 4 support the earlier findings that between-channel gap thresholds are higher than within-channel ones, and that only the strictly between-channel case showed sensitivity to leading-element duration (compare Figs. 3–5). Interestingly, for very short (5 ms) leading-element durations in the between-channel case of experiment 4, the mean gap threshold was 32.3 ms which matches very closely that seen in the between-channel case in which perceptual channels were defined by stimulus spectrum (32.4 ms: experiment 3). The finding of long gap thresholds for diotically presented noises is not completely without precedent. Penner (1977), in studies of decay of sensation, found gap thresholds in the order of 30 ms for diotically presented noise. Moreover, using the present terminology, Penner’s diotic condition was the only one meeting the criteria for “between-channel” gap detection, and it was the only one in that experiment in which gap threshold was independent of the level of the trailing noise burst. This led Penner to conclude that “the diotic detection cue is different from the dichotic” (p. 554).

V. GENERAL DISCUSSION

A. Properties of the between-channel effect

The general purpose of the foregoing experiments was to explore gap-detection performance in normal listeners under conditions in which the mechanisms that mediated the temporal detection task either resided in the central projection of any given cochlear channels (“within-channel” gap detection) or did not (“between-channel” gap detection). In the former case, the perceptual operation required to perform

the task is a simple discontinuity detection in the perceptual channels activated by the stimulus. In the latter case, the detection of the gap must reflect a central computation, i.e., a relative timing of the offset of activity in the channels subserving the leading element and the onset of activity in the channels mediating the trailing element. Each experiment in the present study revealed consistently longer gap thresholds in the between-channel case, suggesting that the mechanisms or processes mediating the two forms of gap-detection performance must have different temporal acuities. Experiments 2 and 4 revealed that only in the between-channel case was there a vulnerability of gap-detection thresholds to the duration of the leading stimulus element. Experiment 3 provided a further example of this effect.

Experiment 1 revealed that the introduction of a spectral disparity between the noise markers comprising the leading and trailing edges of a gap results in impoverished gap-detection performance. These observations are compatible with earlier findings using gap-detection and gap discrimination tasks (Divenyi and Danner, 1977; Fitzgibbons *et al.*, 1974; Formby and Forrest, 1991; Neff *et al.*, 1982) and extend them in two ways. First, the effect extended over a range of at least two octaves in every listener, at which point the effect size was sometimes close to an order of magnitude, i.e., minimal detectable gaps were sometimes nearly ten times poorer than those seen in the traditional, within-channel case in the same subjects. In this regard, the within-channel performance of the listeners in this study was as good as, or better than, other published accounts.

Second, the present data offer some evidence of an asymmetry in this effect: Three of five subjects showed evidence of a plateau effect for stimuli in which the trailing noise marker was more than an octave above the leading one, and four of the five subjects had better gap thresholds for stimuli with higher-frequency trailing markers than for lower-frequency trailing markers. The tendency towards a plateau effect for higher-frequency trailing stimuli offers modest support for the notion of a two-process model: gap detection within a frequency-specific processing channel being mediated by one mechanism, and between-channel gap detection being mediated by a different one. Formby *et al.* (1996) observed a similar asymmetry in studies using sinusoidal markers (see also Heinz *et al.*, 1996).

The further question is why plateau effects of the same kind were less obvious (or not seen at all) on the low-frequency sides of the functions in Fig. 2. One possibility is that the very low-frequency auditory channels (significantly less than 500 Hz) have inherently poorer temporal resolution, because of stimulus uncertainty (i.e., inherent fluctuations in the low-frequency stimulus envelope that might be confused with an inserted gap), or because of some imprecision in central integration or decision processes (after Plack and Moore, 1990; see also Phillips *et al.*, 1994a; Stuart *et al.*, 1995). That is, there may be an inherently less precise “time stamping” of (i.e., less precise allocation of an event time to) auditory events in the lowest frequency channels, and this effect is superimposed on what may otherwise be the two-process model described above.

B. Mechanisms underlying the between-channel effect

Following Fitzgibbons *et al.* (1974) [see also Jones and Yee (1993)], we posit a role for auditory attentional processes in the detection of gaps in the between-channel case, but we prefer to keep the explanation in low-level perceptual rather than cognitive terms. Specifically, we suggest that allocation of perceptual or attentive resources to one perceptual channel (and we emphasize that by “channel,” we mean the entire network of peripheral and central auditory mechanisms activated by the stimulus, and not simply a single critical band) results in an imprecise time-stamping of acoustic events in the nonattended channels. Thus, allocation of attention or perceptual resources to the frequency of the leading element impoverishes the listener’s ability to make precise judgments about the relative timing of the onset of the trailing element. This is not necessarily because of any inherent temporal imprecision within a particular perceptual channel (after Plack and Moore, 1990), but because the selective attention process reduces the resources available for the time-stamping of events in the nonattended channel. Increasing the frequency disparity between the leading and trailing elements in the gap-detection stimulus impoverishes gap-detection performance to the extent that the perceptual timing operation must increasingly rely on a cross correlation of response event times in different channels rather than a discontinuity detection within the attended channel. We know independently that, at threshold levels of stimulation, selective attention in the frequency domain can have profound effects on the *detection* of events in nonattended channels (Scharf *et al.*, 1994; see also Wright and Dai, 1994, for evidence that “channel” here need not refer to a single peripheral auditory filter). The present argument merely extends the effects of selective attention to include the time-stamping of auditory events, even at suprathreshold stimulus levels, in nonattended channels.

A further question is whether the data from experiment 1 have clear implications either for the “tuning” of a central, frequency-specific processing channel, or for the temporal properties of any attentional shift mechanism. Certainly, it is tempting to speculate that an attentional shift mechanism operates on the tonotopic central auditory maps (which are often thought of as “maps” of stimulus frequency), and that increases in the “neuroanatomical distance” to be spanned by the shift determines the duration of the shift process, and thus the shortest gap that can be detected (after the general argument of Divenyi and Danner, 1977). The neurophysiological reality, however, is that tonotopy in the central auditory system is a strictly threshold phenomenon: The *representation* of sound frequency, understood as the spatial distribution of neurons activated by a tonal stimulus, can span two to three octaves of the tonotopic map at suprathreshold levels, and in a strikingly patchy, discontinuous fashion (Phillips *et al.*, 1994b). This makes the concepts of both the “tuning” of a central, frequency-specific processing channel and “neuroanatomical distance” difficult ones to formalize. One might retreat from the neuroscience and conceptualize the attentional shift as being across subjective

pitch space, rather than across a tonotopic map, but the implications of such a stance are less than immediately clear.

ACKNOWLEDGMENTS

This work was supported by NSERC Research Grants to DPP and NSERC and Killam awards to TLT. Special thanks are due to Dr. A. Stuart and Dr. R. M. Klein for helpful comments on earlier presentations of this work, and to Dr. C. Formby and an anonymous reviewer for their commentaries on previous versions of this manuscript.

- Buchtel, H. A., and Stewart, J. D. (1989). “Auditory agnosia: A perceptive or associative disorder?” *Brain Lang.* **37**, 12–25.
- Carney, L. H., and Geisler, C. D. (1986). “A temporal analysis of auditory-nerve fiber responses to spoken stop consonant-vowel syllables,” *J. Acoust. Soc. Am.* **79**, 1896–1914.
- Divenyi, P. L., and Danner, W. F. (1977). “Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra,” *Percept. Psychophys.* **21**, 125–142.
- Eddins, D. A., Hall, J. W., and Grose, J. H. (1992). “The detection of temporal gaps as a function of frequency region and absolute noise bandwidth,” *J. Acoust. Soc. Am.* **91**, 1069–1077.
- Eggermont, J. J. (1995). “Neural correlates of gap detection and auditory fusion in cat auditory cortex,” *Neuroreport* **6**, 1645–1648.
- Eimas, P. D., and Corbit, J. D. (1973). “Selective adaptation of linguistic feature detectors,” *Cogn. Psychol.* **4**, 99–109.
- Fitzgibbons, P. J. (1983). “Temporal gap detection in noise as a function of frequency, bandwidth, and level,” *J. Acoust. Soc. Am.* **72**, 761–765.
- Fitzgibbons, P. J., Pollatsek, A., and Thomas, I. B. (1974). “Detection of temporal gaps within and between perceptual tonal groups,” *Percept. Psychophys.* **16**, 522–528.
- Fitzgibbons, P. J., and Wightman, F. L. (1982). “Gap detection in normal and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **74**, 761–765.
- Florentine, M., and Buus, S. (1984). “Temporal gap detection in sensorineural and simulated hearing impairments,” *J. Speech Hear. Res.* **27**, 449–455.
- Formby, C., Barker, C., Abbey, H., and Raney, J. J. (1993). “Detection of silent temporal gaps between narrow-band noise markers having second-formantlike properties of voiceless stop/vowel combinations,” *J. Acoust. Soc. Am.* **93**, 1023–1027.
- Formby, C., and Forrest, T. G. (1991). “Detection of silent temporal gaps in sinusoidal markers,” *J. Acoust. Soc. Am.* **89**, 830–837.
- Formby, C., and Muir, K. (1988). “Modulation and gap detection for broadband and filtered noise signals,” *J. Acoust. Soc. Am.* **84**, 545–550.
- Formby, C., Sherlock, L. P., and Forrest, T. G. (1996). “An asymmetric roex filter model for describing detection of silent gaps in sinusoidal markers,” *Aud. Neurosci.* **3**, 1–20.
- Forrest, T. G., and Formby, C. (1996). “Detection of silent temporal gaps in sinusoidal markers simulated with a single-channel envelope detector model,” *Aud. Neurosci.* **3**, 21–33.
- Green, D. M., and Forrest, T. G. (1989). “Temporal gaps in noise and sinusoids,” *J. Acoust. Soc. Am.* **86**, 961–970.
- Grose, J. H. (1991). “Gap detection in multiple narrow bands of noise as a function of spectral configuration,” *J. Acoust. Soc. Am.* **90**, 3061–3068.
- Grose, J. H., Eddins, D. A., and Hall, J. W. (1989). “Gap detection as a function of stimulus bandwidth with fixed high-frequency cutoff in normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **86**, 1747–1755.
- Hall, J. W., Grose, J. H., and Joy, S. (1996). “Gap detection for pairs of noise bands: Effects of stimulus level and frequency separation,” *J. Acoust. Soc. Am.* **99**, 1091–1095.
- Heinz, M. G., Goldstein, M. H., Jr., and Formby, C. (1996). “Temporal gap detection thresholds in sinusoidal markers simulated with a multi-channel, multi-resolution model of the auditory periphery,” *Aud. Neurosci.* **3**, 35–56.
- Ison, J. R., O’Conner, K., Bowen, G. P., and Bocirnea, A. (1991). “Temporal resolution of gaps in noise by the rat is lost with functional decortication,” *Behav. Neurosci.* **105**, 33–40.

- Jones, M. R., and Yee, W. (1993). "Attending to auditory events: the role of temporal organization," in *Thinking in Sound. The Cognitive Psychology of Human Hearing*, edited by S. McAdams and E. Bigand (Oxford U.P., Oxford, England), pp. 69–112.
- Kelly, J. B., Rooney, B. J., and Phillips, D. P. (1996). "Effects of bilateral auditory cortical lesions on gap detection thresholds in the ferret (*Mustela putorius*)," *Behav. Neurosci.* **110**, 542–550.
- Kuhl, P. K., and Miller, J. D. (1978). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *J. Acoust. Soc. Am.* **63**, 905–917.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Moore, B. C. J. (1985). "Frequency selectivity and temporal resolution in normal and hearing-impaired listeners," *Br. J. Audiol.* **19**, 189–201.
- Moore, B. C. J., and Glasberg, B. R. (1988). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears," *J. Acoust. Soc. Am.* **83**, 1093–1101.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1993). "Detection of temporal gaps in sinusoids: Effects of frequency and level," *J. Acoust. Soc. Am.* **93**, 1563–1570.
- Morse, P. A., and Snowden, C. T. (1975). "An investigation of categorical speech discrimination by rhesus monkeys," *Percept. Psychophys.* **17**, 9–16.
- Neff, D. L., Jesteadt, W., and Brown, E. L. (1982). "The relation between gap discrimination and auditory stream segregation," *Percept. Psychophys.* **31**, 493–501.
- Nelson, P. B., Nitttrouer, S., and Norton, S. J. (1995). "'Say-stay' identification and psychoacoustic performance of hearing-impaired listeners," *J. Acoust. Soc. Am.* **97**, 1830–1838.
- Penner, M. J. (1977). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," *J. Acoust. Soc. Am.* **61**, 552–557.
- Phillips, D. P., Rappaport, J. M., and Gulliver, J. M. (1994a). "Impaired word recognition in noise by patients with noise-induced cochlear hearing loss: contribution of temporal resolution defect," *Am. J. Otol.* **15**, 679–686.
- Phillips, D. P., Semple, M. N., Calford, M. B., and Kitzes, L. M. (1994b). "Level-dependent representation of stimulus frequency in cat primary auditory cortex," *Exp. Brain Res.* **102**, 210–226.
- Plack, C. J., and Moore, B. C. J. (1990). "Temporal window shape as a function of frequency and level," *J. Acoust. Soc. Am.* **87**, 2178–2187.
- Plomp, R. (1964). "Rate of decay of auditory sensation," *J. Acoust. Soc. Am.* **36**, 277–282.
- Scharf, B. (1970). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.
- Scharf, B., Magnan, J., Collet, L., Ulmer, E., and Chays, A. (1994). "On the role of the olivocochlear bundle in hearing: a case study," *Hearing Res.* **75**, 11–26.
- Shailer, M. J., and Moore, B. C. J. (1983). "Gap detection as a function of frequency, bandwidth, and level," *J. Acoust. Soc. Am.* **74**, 467–473.
- Sinex, D. G., and McDonald, L. P. (1988). "Average discharge rate representation of voice onset time in chinchilla auditory nerve," *J. Acoust. Soc. Am.* **83**, 1817–1827.
- Sinex, D. G., and Narayan, S. S. (1994). "Auditory-nerve fiber representation of temporal cues to voicing in word-medial stop consonants," *J. Acoust. Soc. Am.* **95**, 897–903.
- Stevens, K. N., and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants," *J. Acoust. Soc. Am.* **64**, 1358–1368.
- Stuart, A., Phillips, D. P., and Green, W. B. (1995). "Word recognition performance in continuous and interrupted broad-band noise by normal-hearing and simulated hearing-impaired listeners," *Am. J. Otol.* **16**, 658–663.
- Williams, K. N., and Perrott, D. R. (1972). "Temporal resolution of tonal pulses," *J. Acoust. Soc. Am.* **51**, 644–647.
- Wright, B. A., and Dai, H. (1994). "Detection of unexpected tones with short and long durations," *J. Acoust. Soc. Am.* **95**, 931–938.
- Zhang, W., Salvi, R. J., and Saunders, S. S. (1990). "Neural correlates of gap detection in auditory nerve fibers of the chinchilla," *Hearing Res.* **46**, 181–200.

Psychometric functions and temporal integration in electric hearing

Gail S. Donaldson,^{a)} Neal F. Viemeister, and David A. Nelson

Departments of Otolaryngology and Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 20 June 1996; revised 18 July 1996; accepted 17 January 1997)

Temporal-integration functions and psychometric functions for detection were obtained in eight users of the Nucleus 22-electrode cochlear implant. Stimuli were 100-Hz, 200- μ s/phase trains of biphasic pulses with durations ranging from 0.44 to 630.44 ms (1 to 64 pulses). Temporal-integration functions were measured for 21 electrodes. Slopes of these functions were considerably shallower than the 2.5 dB/doubling slopes typically observed in acoustic hearing. They varied widely across subjects and for different electrodes in a given subject, ranging from 0.06 to 1.94 dB/doubling of stimulus pulses, with a mean [standard deviation (s.d.)] value of 0.42 (0.38). Psychometric functions were measured for 11 of the same 21 electrodes. Slopes of psychometric functions also varied across subjects and electrodes, and were 2–20 times steeper than those reported by other investigators for normal-hearing and cochlear-impaired acoustic listeners. Slopes of individual psychometric functions for 1-, 2-, 4-, and 8-pulse stimuli ranged from 0.20 to 1.84 log d' /dB with a mean (s.d.) value of 0.77 (0.45). Psychometric-function slopes did not vary systematically with stimulus duration in most cases. A clear inverse relation between slopes of psychometric functions and slopes of temporal-integration functions was observed. This relation was reasonably well described by a hyperbolic function predicted by the multiple-looks model of temporal integration [Viemeister and Wakefield, *J. Acoust. Soc. Am.* **90**, 858–865 (1991)]. Psychometric-function slopes tended to increase with absolute threshold and were inversely correlated with dynamic range, suggesting that observed differences in psychometric-function slopes across subjects and electrodes may reflect underlying differences in neural survival. © 1997 Acoustical Society of America. [S0001-4966(97)03506-6]

PACS numbers: 43.66.Ts, 43.66.Cb, 43.66.Mk, 43.64.Me [JWH]

INTRODUCTION

An important phenomenon observed in hearing and other sensory systems is that detection thresholds for brief stimuli improve with increasing stimulus duration. This phenomenon, termed temporal integration, has been described in audition for both normal-hearing and hearing-impaired listeners, for a variety of tonal and broadband stimuli [see Gerken *et al.* (1990), for a review]. Normal-hearing listeners demonstrate approximately 2.5 dB threshold improvement per doubling of stimulus duration for durations up to 200–300 ms. Temporal integration is reduced in the presence of cochlear hearing loss, with slopes shallower than 1 dB/doubling often observed (e.g., Gengel and Watson, 1971; Florentine *et al.*, 1988).

Classic models of temporal integration assume “leaky” (Plomp and Bouman, 1959; Zwislocki, 1960, 1969; Jeffress, 1967, 1968) or perfect (Green, 1960; Green and Swets, 1966) integration of stimulus energy with a time constant on the order of several hundred milliseconds. These models can account satisfactorily for the shapes of temporal-integration functions, but their long time constants grossly exceed the integration time constants necessary to explain temporal-resolution phenomena such as modulation detection and gap detection. Other models, that utilize shorter time constants

(<10 ms), are able to account for various temporal-resolution data (Viemeister, 1979; Forrest and Green, 1987; Moore *et al.*, 1988); however, their predictions of temporal-integration functions at threshold are unrealistic. Thus, the assumption of a single time constant for integration of stimulus energy, whether short or long, appears unable to explain both temporal-integration and temporal-resolution phenomena. This dichotomy has been termed the “integration-resolution paradox” (Green, 1985; de Boer, 1975).

A solution to this paradox was proposed by Viemeister and Wakefield (1991), who reported data from two experiments suggesting that long time-constant integration does not underlie temporal integration but, instead, that multiple, short “looks” at the processed signal are combined to improve detection performance. An overview of the *multiple-looks* model is as follows: Signals within a single channel (e.g., critical-band filter) undergo a nonlinear transformation such as half-wave rectification, followed by leaky integration within a temporal window having a short time constant (~ 3 ms). The ongoing output of the short temporal window is stored in short-term memory. Multiple samples of the processed signal are available in short-term memory for use by the observer in making a decision related to signal detection or discrimination. In its simplest implementation, considered here, the samples or “looks” are assumed to be independent, optimally combined, and equally detectable. Performance (d') at a given stimulus intensity improves with n , the number of looks, according to the relation: $d'_n = \sqrt{n}d'_1$. In the

^{a)}Correspondence to: Gail S. Donaldson, Ph.D., Box 396 UMHC, Room 8-323 PWB, 516 Delaware St., S.E., Minneapolis, MN 55455. Electronic mail: donal005@maroon.tc.umn.edu

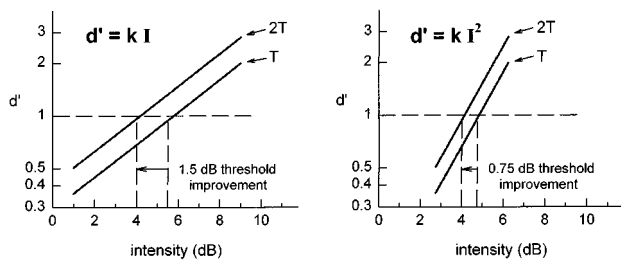


FIG. 1. Illustration of the relationship between psychometric function slope and predicted threshold improvement. In each panel, lower and upper solid lines represent psychometric functions for stimuli with durations of T and $2T$, respectively. Functions shown in the left panel are shallower (slopes of $1.0 \log d'/\text{dB}$) than those in the right panel (slopes of $2.0 \log d'/\text{dB}$). According to the simple version of the multiple-looks model, a doubling of stimulus duration improves performance (d') by $\sqrt{2}$ at all stimulus levels, resulting in an upward shift of the psychometric function. For a particular threshold criterion (e.g., $d' = 1$, shown by the dashed horizontal line), improved performance corresponds to a decrease in threshold; however, the magnitude of threshold improvement is inversely proportional to the slope of the psychometric function. In this example, doubling stimulus duration results in 1.5-dB threshold improvement in the case of the shallower psychometric function (left panel), but only 0.75-dB threshold improvement in the case of the steeper function (right panel).

case of temporal integration, an increase in stimulus duration corresponds to an increase in the number of available looks; this results in improved performance at a given stimulus intensity or, for fixed performance, a reduction in “threshold.”

An important prediction of the multiple-looks model of temporal integration is that the amount of threshold improvement associated with an increase in the number of looks depends on the slope of the psychometric function for a single look. As demonstrated in Fig. 1, steep psychometric functions will result in less threshold improvement than shallow psychometric functions, because the intensity difference corresponding to a constant (\sqrt{n}) improvement in performance (d') will be smaller in the case of a steeper function.¹ As noted by Viemeister and Wakefield (1991), this aspect of the theory is consistent, at least qualitatively, with the reduced temporal integration observed in cochlear hearing loss and with the dependence of temporal-integration slopes on stimulus frequency in normal-hearing listeners. However, direct evaluation of the dependence of temporal-integration slope on the slope of the psychometric function for a single look has proven to be difficult because the range of temporal-integration slopes observed in acoustic listeners is

small and because the resolution with which psychometric-function slopes can be estimated is limited by normal sources of variability.

Existing reports of temporal integration in electric hearing indicate that slopes of temporal-integration functions vary considerably across subjects and for different stimulus conditions (Eddington *et al.*, 1978; Shannon, 1983, 1986, 1990; White, 1984; Pfungst *et al.*, 1991; Moon *et al.*, 1993; Pfungst and Morris, 1993). Given this, and assuming that the multiple-looks model is correct, then psychometric-function slopes for electric stimulation should also vary. Pfungst *et al.* (1991) considered the relation between psychometric-function slopes and temporal-integration slopes for behavioral data from implanted monkeys; however, their analyses appear to have been limited to qualitative observations, and these did not reveal any clear relation between the two measures. To our knowledge, psychometric functions have not been described in human listeners with cochlear implants.

The present study was undertaken with two goals in mind: One goal was to provide detailed descriptions of psychometric functions for detection and corresponding temporal-integration functions in a group of adult cochlear-implant users. The second goal was to determine whether psychometric-function slopes are predictive of temporal-integration function slopes in electric hearing, as suggested by the multiple-looks hypothesis.

I. METHODS

A. Subjects

Eight adult users of the Nucleus 22-electrode cochlear implant served as subjects. Onset of deafness was postlingual in seven subjects and prelingual in the remaining subject (AMA). Table I lists each subject's primary cause of deafness, years of deafness prior to implant surgery, age at implantation, depth of electrode-array insertion, and years of implant use at the time of data collection. As indicated, five of eight subjects were long-term users of the implant (5–7 years) at the time of testing. The other three (AMB, CXL, and SLB) had one year or less experience with their devices, but obtained considerable benefit from them in everyday use.

Experiments involved the presentation of electric stimuli on bipolar electrode pairs. Temporal-integration functions were obtained for 21 electrodes (1–3 electrodes in each subject); psychometric functions for detection were obtained for

TABLE I. Description of eight cochlear-implant subjects who participated in the present study. Subject identifying code, primary cause of deafness, years of deafness (implanted ear) prior to implantation, age at implantation, depth of electrode-array insertion (mm from the round window, with 25 mm representing complete insertion), and years of implant use at the time of the present testing are listed for each subject.

| Subject | Cause of deafness | Years deaf | Age | Depth | Years use |
|---------|---------------------------|------------|-----|-------|-----------|
| AMA | Congenital | 57 | 57 | 19.4 | 6 |
| AMB | Progressive SNHL | 1 | 49 | 25.0 | 1 |
| CXL | Aminoglycoside toxicity | 1.5 | 32 | 24.0 | 0.5 |
| EES | Cogan's syndrome | 4 | 54 | 17.0 | 7 |
| FXC | Progressive SNHL | 4 | 64 | 25.0 | 5 |
| JPB | Progressive SNHL | 4 | 52 | 24.0 | 5 |
| SLB | Progressive SNHL | 10 | 47 | 25.0 | 1 |
| VVK | Sudden, After ear surgery | 38 | 55 | 17.0 | 7 |

TABLE II. Detection threshold (THS), maximum acceptable loudness level (MAL) and dynamic range (DR) for each of 21 test electrodes in eight cochlear-implant subjects. Stimuli were 500-ms trains of 125-Hz, 205- μ s/phase, biphasic pulses. Data represent means of values obtained at one or more timepoints within six months of data collection for the present study.

| Subject | Electrode | THS (dB re: 1 μ A) | MAL (dB re: 1 μ A) | DR (dB) |
|---------|-----------|---------------------------|---------------------------|------------|
| AMA | rEL04 | 57.3 | 61.4 | 4.1 |
| | rEL19 | 53.5 | 56.7 | 3.2 |
| AMB | rEL06 | 48.1 | 56.0 | 8.0 |
| | rEL12 | 47.1 | 56.6 | 9.5 |
| | rEL20 | 47.6 | 57.6 | 10.0 |
| | rEL05 | 48.8 | 64.8 ^a | > 16.0 |
| CXL | rEL11 | 45.8 | 64.8 ^a | > 18.0 |
| | rEL18 | 44.4 | 64.8 ^a | > 20.4 |
| | rEL05 | 49.2 | 52.6 | 3.6 |
| EES | rEL11 | 49.6 | 53.8 | 3.9 |
| | rEL18 | 48.6 | 52.7 | 4.0 |
| | rEL05 | 49.3 | 55.2 | 6.0 |
| FXC | rEL11 | 50.9 | 57.7 | 6.7 |
| | rEL18 | 52.3 | 55.9 | 3.6 |
| | rEL07 | 51.7 | 57.8 | 6.1 |
| JPB | rEL11 | 51.0 | 57.8 | 6.9 |
| | rEL16 | 53.1 | 60.5 | 7.4 |
| | rEL21 | 53.6 | 57.6 | 4.1 |
| SLB | rEL06 | 48.7 | 53.2 | 4.5 |
| VVK | rEL10 | 47.4 | 50.9 | 3.5 |
| | rEL15 | 49.1 | 50.7 | 1.6 |

^aMAL not reached at current–amplitude limits of receiver–stimulator.

11 of these electrodes. The spatial separation between active and reference electrodes was set to agree with that used in subjects' speech-processor maps. This separation was 1.5 mm (7 subjects) or 0.75 mm (subject AMA). In this paper, research electrode (rEL) numbering is used, i.e., electrodes are numbered consecutively from the apical to basal end of the 22-electrode array, and electrode number refers to the basal member of the stimulating pair.

Electrodes were chosen to represent a range of absolute thresholds and dynamic ranges, and different regions of subjects' implanted arrays. Specifically, subjects were tested whose average dynamic ranges varied from narrow (<4 dB) to relatively wide (>8 dB) and, to the extent possible, test electrodes in a given subject were chosen to represent differing thresholds and dynamic ranges. Table II lists detection threshold (THS), maximum acceptable loudness level (MAL), and dynamic range (DR) for each electrode. These measures were obtained in response to a standard stimulus (500-ms train of 125-Hz, 205- μ s/phase biphasic pulses) using an ascending method of adjustment. In the adjustment procedure, pulse trains are repeated continuously at a rate of 1/s and the experimenter slowly increases stimulus current amplitude until the subject indicates that the sound is just audible. This level is recorded and then remeasured in another ascending run. The mean of two or three such estimates is taken to be threshold. An estimate of MAL is subsequently obtained by increasing stimulus level above the threshold estimate until the subject indicates that stimulus loudness has reached a level that can only be tolerated for a short time. THS and MAL estimates for this standard stimulus are obtained periodically in our laboratory as part of a baseline characterization of subjects' electric hearing and to

monitor for obvious changes in hearing sensitivity over time. Values in Table II represent mean data for each subject from one or more test dates over a time period within ± 6 months of data collection for the present experiments. They are presented here to provide the reader with an overview of the sensitivity and dynamic range characteristics of electrodes tested. Additional measures of THS and MAL were obtained as part of the present experiments, and are described below.

B. Stimuli

Stimuli for the present experiments were pulse trains comprised of 1, 2, 4, 8, 16, 32, or 64 biphasic pulses presented at a rate of 100 pulses/s. Individual pulses had a fixed pulse duration of 200 μ s/phase, with a 44- μ s interphase gap; thus, interpulse interval for the multiple-pulse stimuli was 9.56 ms, and train durations for the 1- to 64-pulse stimuli ranged from 0.44 to 630.44 ms. Current amplitude (pulse height) was always the parameter varied.

Experiments were controlled by a 50-MHz 80486 computer connected through a parallel port to a BTNI cochlear-implant interface (Shannon *et al.*, 1990). Stimulus amplitudes called for in the adaptive and fixed-level procedures (described below) were translated to the nearest current value that could be delivered by a particular subject's receiver–stimulator. These values corresponded to integer *current step units* (CSUs),² and were determined using calibration tables provided by Cochlear Corporation for each subject's implanted device. Stimuli were verified using a calibrated, Nucleus 22-electrode receiver–stimulator whose outputs could be monitored directly (*Implant-in-a-Box*, Cochlear Corporation).

C. Psychophysical procedures

Psychophysical procedures were selected that would permit collection of an entire temporal-integration function or psychometric function for detection in a single testing session. This was necessary to ensure that small shifts in subjects' hearing sensitivity across sessions did not invalidate estimates of temporal-integration slope or psychometric-function slope. Threshold data used to construct temporal-integration functions were obtained using an adaptive procedure; psychometric functions were obtained with a fixed-level procedure. Details of these procedures are given below.

Temporal-integration functions. Temporal-integration functions were constructed from threshold data for each of the seven stimuli. For each stimulus, preliminary estimates of threshold and maximum acceptable loudness were obtained using the ascending method of adjustment procedure described earlier (Sec. I A). These values were used to set the starting level and maximum-safe-stimulation level, respectively, for a 3IFC adaptive procedure that provided final threshold estimates comprising temporal-integration functions. In the adaptive procedure, three listening intervals were cued visually on a video monitor and the stimulus was presented in one of the intervals, chosen at random on each trial. The subject's task was to identify the interval contain-

ing the stimulus and to press the corresponding button on a three-button computer mouse. Correct-answer feedback was provided after each trial.

Stimulus level was set 1–3 dB above the preliminary (adjustment) threshold estimate at the start of each adaptive track. The first four reversals were obtained using a step size of either 2 or 4 CSUs (approximately 0.3 or 0.6 dB) and a two-down, one-up stepping rule. These initial trials were intended to quickly move the adaptive track to a level near detection threshold. Following the fourth reversal, step size was halved and a three-down, one-up stepping rule was assumed. This rule estimates the stimulus level necessary to obtain 79.4% correct responses (Levitt, 1971). Trials continued until eight additional reversals occurred. Threshold was taken to be the arithmetic mean of the current levels (dB *re*: 1 μ A) at the final eight reversals.

Threshold estimates were obtained in “sets,” where a set consisted of one threshold estimate for each of the seven stimuli (1, 2, 4, 8, 16, 32, or 64 pulses), collected in order of increasing stimulus duration.³ Three to four complete sets were obtained initially. Additional tracks were then obtained, if needed, to replace tracks in which reversals failed to converge or to provide additional threshold estimates if the first 3 to 4 estimates for a particular stimulus were unusually variable.⁴ Temporal-integration functions were constructed from means of 3–5 threshold estimates for each stimulus. Data comprising a given function were typically obtained in a single 2-h session; however, it was occasionally necessary to combine data across two sessions. When this was the case, threshold data collected during a second session were evaluated for shifts in the subject’s sensitivity. If such a shift was suspected, the original threshold estimates were discarded and another complete set of data was obtained.

Psychometric functions for detection. Psychometric functions for detection were obtained for the 1-, 2-, 4-, 8-, and 64-pulse stimuli, using a fixed-level, 2IFC procedure. In this procedure, two listening intervals were cued on a video monitor. The stimulus was presented in one of the intervals, chosen at random on each trial. Again, correct-answer feedback was provided after each trial.

Stimulus levels for a particular psychometric function were selected to span the range of near-chance to near-perfect performance in 5–10 steps. Levels could not be specified uniformly for all electrodes and stimuli (e.g., at fixed levels relative to the adaptive threshold), first, because the rate of performance increase with level varied across subjects and, second, because subjects’ receiver–stimulators could deliver current levels only in discrete steps corresponding to integer CSUs. Instead, for a given psychometric function, the interval between adjacent stimulus levels was held constant at 1, 2, or 4 CSUs (approximately 0.15, 0.3, or 0.6 dB, respectively). Appropriate stimulus levels were initially estimated from the 3IFC threshold and were modified, if necessary, after an initial descending run was completed. Data were collected in blocks of 20 trials. The first data set consisted of one block of trials at each stimulus level, in descending order. Stimulus level order was alternated between ascending and descending in subsequent sets.⁵ Four to six data sets (80–120 trials per point) were obtained in this

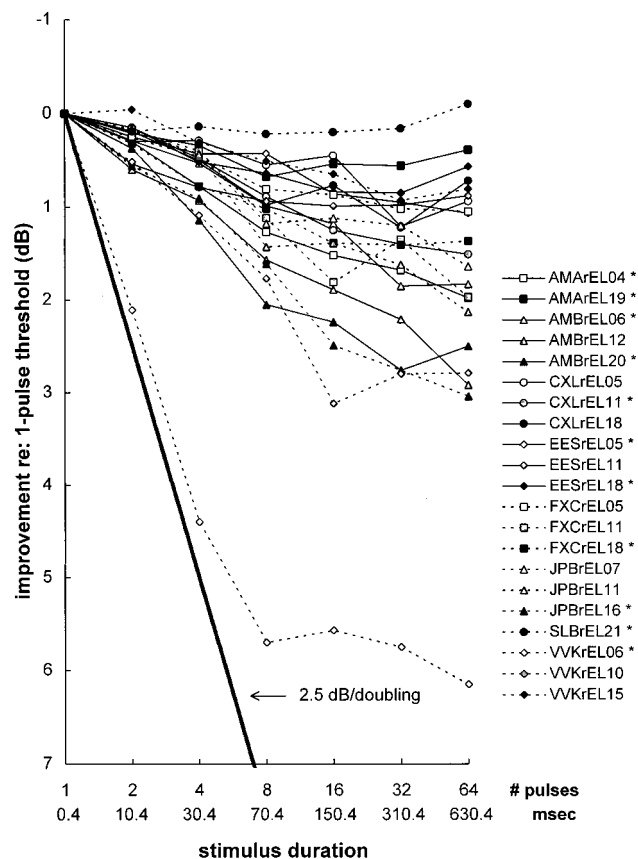


FIG. 2. Temporal-integration functions for 21 test electrodes in eight cochlear-implant subjects. All thresholds are expressed relative to the threshold for a single pulse in order to facilitate comparison of function shapes across electrodes. Different combinations of symbols and connecting lines represent different electrodes, as indicated in the legend; data from apical, middle, and basal electrodes are plotted as open, shaded, and filled symbols, respectively. Asterisks next to legend entries indicate electrodes for which psychometric functions were also obtained (see Fig. 6). The heavy, solid line represents the slope (2.5 dB/doubling of pulses) expected for normal-hearing listeners performing a similar, acoustic task.

manner, with all data for a particular psychometric function collected in a single, 2 to 3 h session.

Slope estimates for each psychometric function were obtained as follows: The mean percent-correct score and corresponding standard deviation were computed for the 4–6 blocks of trials obtained at each stimulus level, and the 90% confidence interval (CI) around each mean value was determined. Data points were included in psychometric-function fits if the lower limit of the CI was greater than 50% and the upper limit of the CI was less than 100%.⁶ Mean percent-correct values were transformed to d' units, and a linear function was fit to the data in $\log d'$ versus dB coordinates by least-squares regression.

II. RESULTS AND DISCUSSION

A. Temporal-integration functions

Variability of temporal-integration slopes across subjects and electrodes. Temporal-integration functions for all 21 test electrodes are shown in Fig. 2. To facilitate comparison of function shapes across electrodes, thresholds for each

TABLE III. Slopes of short-duration segments of temporal-integration functions (1–8 pulses), estimated by linear regression of threshold (dB re: 1 μ A) on stimulus duration (doublings of number pulses). Standard errors (se) of slope estimates and 90% confidence intervals (90% CI) are also given. Slopes are expressed as positive numbers (i.e., dB threshold improvement per doubling of number pulses). With one exception (SLB rEL12), significant r^2 values ($r^2 > 0.85$, $p < 0.01$) were associated with all slope estimates.

| Subject | Electrode | Slope | se | 90% CI |
|---------|-----------|-------|-------|------------|
| AMA | rEL04 | 0.43 | 0.032 | 0.36–0.52 |
| | rEL19 | 0.22 | 0.029 | 0.13–0.30 |
| AMB | rEL06 | 0.32 | 0.033 | 0.23–0.42 |
| | rEL12 | 0.50 | 0.042 | 0.38–0.63 |
| | rEL20 | 0.69 | 0.087 | 0.44–0.95 |
| CXL | rEL05 | 0.16 | 0.042 | 0.04–0.29 |
| | rEL11 | 0.30 | 0.042 | 0.18–0.42 |
| | rEL18 | 0.33 | 0.046 | 0.19–0.46 |
| EES | rEL05 | 0.15 | 0.040 | 0.03–0.27 |
| | rEL11 | 0.31 | 0.063 | 0.12–0.49 |
| | rEL18 | 0.23 | 0.033 | 0.13–0.33 |
| FXC | rEL05 | 0.26 | 0.021 | 0.20–0.32 |
| | rEL11 | 0.36 | 0.092 | 0.09–0.63 |
| | rEL18 | 0.35 | 0.028 | 0.27–0.44 |
| JPB | rEL07 | 0.38 | 0.095 | 0.10–0.66 |
| | rEL11 | 0.46 | 0.137 | 0.06–0.86 |
| | rEL16 | 0.52 | 0.045 | 0.39–0.65 |
| SLB | rEL21 | 0.06 | 0.031 | –0.03–0.15 |
| VVK | rEL06 | 1.94 | 0.152 | 1.49–2.38 |
| | rEL10 | 0.59 | 0.026 | 0.51–0.66 |
| | rEL15 | 0.19 | 0.054 | 0.03–0.35 |

electrode are expressed relative to the threshold for a single pulse presented to the same electrode. The heavy, solid line in Fig. 2 has a slope of 2.5 dB/doubling; this corresponds to the slope of a temporal-integration function expected from a normal-hearing listener performing a similar, acoustic task (Gerken *et al.*, 1990; Carlyon *et al.*, 1990).

Figure 2 shows that the slopes of temporal-integration functions and, related to this, the maximum threshold improvement obtained with increasing stimulus duration, varied considerably across subjects and electrodes. At one extreme, the function obtained for electrode SLBrEL21 (topmost curve in Fig. 2) showed essentially no temporal integration. At the opposite extreme, electrode VVKrEL06 (bottom-most curve in Fig. 2) demonstrated nearly 6 dB of threshold improvement as the number of stimulus pulses increased from 1 to 8, with the temporal-integration slope for short-duration stimuli (1.94 dB/doubling) approaching the value expected in normal-hearing acoustic subjects (2.5 dB/doubling). Slopes of temporal-integration functions' initial segments were quantified by fitting a line to the 1-, 2-, 4-, and 8-pulse data by least-squares regression.⁷ The resulting parameters are given in Table III. Slopes ranged from 0.06 to 1.94 dB threshold improvement per doubling of the number of pulses, with mean and median values of 0.42 and 0.33 dB/doubling, respectively. The shallowest of these slopes (0.06, electrode SLBrEL21) was not significantly different from zero ($p > 0.1$).

The same temporal-integration functions are replotted in Fig. 3, with each panel showing data for a single subject. Mean detection thresholds are shown as symbols connected by thin solid lines; linear functions fit to the 1-, 2-, 4-, and

8-pulse data (described above) are shown as heavy line segments.

Several features of Fig. 3 are noteworthy. First, it can be seen that the variability of threshold estimates, as represented by the size of error bars, was generally small. Electrode VVKrEL06 was an obvious exception to this, demonstrating high variability for the 2- and 4-pulse thresholds. We can offer no explanation for this variability: Subject VVK had no difficulty performing the adaptive threshold task, and there were no systematic shifts in this electrode's threshold estimates over time. The variability of threshold estimates for 1- and 8-pulse stimuli was considerably smaller; thus, it is unlikely that increased variability of the 2- and 4-pulse thresholds had a substantial effect on the overall shape of the temporal-integration function for this electrode, or on its unusually steep short-duration slope.

A second feature of Fig. 3 involves comparison of temporal-integration function slopes across electrodes for the seven subjects with multiple test electrodes. Subjects CXL, EES, FXC, and JPB exhibited relatively similar temporal-integration slopes on different test electrodes. Mean [standard deviation (s.d.)] slope estimates for these subjects were 0.26 (0.09), 0.23 (0.08), 0.33 (0.05), and 0.45 (0.07) dB/doubling, respectively. Somewhat larger slope differences across electrodes were exhibited by subjects AMA and AMB. AMA's electrodes rEL04 and rEL19 yielded slopes of 0.43 and 0.22, respectively, and the 90% confidence intervals of slope estimates for these electrodes were nonoverlapping (see Table III). AMB's electrodes rEL06, rEL12, and rEL20 yielded slopes of 0.32, 0.50, and 0.69 dB/doubling, respectively. Confidence intervals suggest that the temporal-integration slope for rEL06 was shallower than that for rEL20, but that other pairs of slopes (i.e., rEL06 versus rEL12, and rEL12 versus rEL20) were not significantly different. The largest interelectrode slope differences were exhibited by subject VVK, whose electrode rEL06 produced a temporal-integration function more than three times steeper (1.94 dB/doubling) than any other test electrode. Slopes for VVK's middle and apical electrodes (rEL10 and rEL15) fell within the range observed for other subjects (0.59 and 0.19 dB/doubling, respectively); however, 90% confidence intervals indicate significant slope differences between all three electrodes.

It will be shown later that short-duration temporal-integration slopes can be predicted relatively well by the slopes of corresponding psychometric functions for detection. In turn, it may be argued that psychometric-function slopes are determined by the characteristics of surviving neurons nearest the intracochlear stimulating electrodes. If this is the case, then the large differences in temporal-integration slopes exhibited by subject VVK across electrodes may reflect substantial differences in the characteristics of surviving neurons in different regions of the cochlea. Similar but less dramatic differences may account for the variation in temporal-integration slopes across electrodes in subjects AMA and AMB, and for differences in average temporal-integration slopes across subjects.

A final characteristic exhibited by the temporal-integration functions in Fig. 3 is a marked reduction in slope

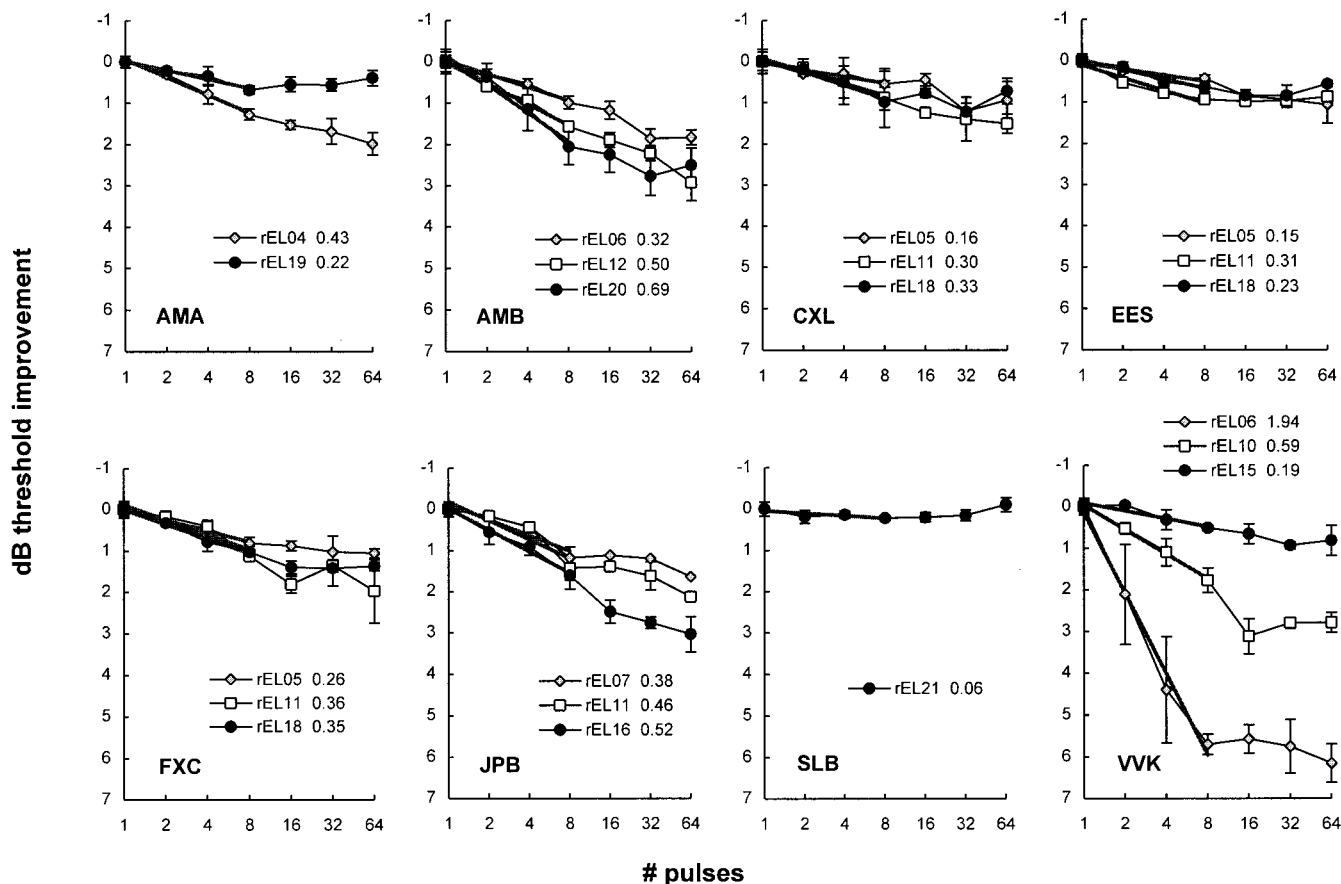


FIG. 3. Temporal-integration functions for 21 electrodes in eight cochlear-implant subjects, replotted from Fig. 2. Each panel shows data for 1–3 electrodes in a single subject whose subject code is given in the panel's lower-left corner. Mean detection thresholds are shown as symbols connected by thin solid lines; error bars represent values ± 1 s.d. from each mean. As in Fig. 2, data for apical, middle, and basal electrodes are plotted with open, shaded, and filled symbols, respectively. Heavy line segments represent linear functions resulting from least-squares regressions to the 1-, 2-, 4-, and 8-pulse data (see text). Electrode numbers and regression slopes are indicated in the symbol key for each panel.

for long-duration stimuli. Such a slope change is particularly evident in the functions for subject VVK's middle and basal electrodes (rEL06 and rEL10): The function for rEL06 possesses a sharp kneepoint at eight pulses, with slope decreasing from 1.94 dB/doubling of pulses for shorter stimuli to a slope that is nearly flat for stimulus trains with eight or more pulses. A similar kneepoint is seen at 16 pulses for rEL10. With one exception (JPB rEL16), least-squares linear-regression estimates of temporal-integration slopes between 16 and 64 pulses were not significantly different from zero ($p > 0.05$). It is not clear what mechanisms underlie the observed flattening of functions at 8–16 pulses.

Threshold change expressed as percent dynamic range.

In Fig. 2, it was shown that the amount of threshold improvement associated with temporal integration is considerably smaller in electric hearing than in acoustic hearing. This finding seems to imply that temporal integration has less influence on the detection of short-duration sounds in the electric case. However, if threshold improvement is expressed as a proportion of dynamic range in both acoustic and electric hearing, a different view is obtained. Consider that the dynamic range of normal, acoustic hearing is roughly 100 dB. Temporal integration at a rate of 2.5 dB per doubling of stimulus duration results in threshold improvements of 10 dB, or about 10% of dynamic range, as signal

duration increases by a factor of 16 (e.g., from 10 to 160 ms). In the electric case, dynamic range may be 5 dB or less. A temporal-integration slope of 0.5 dB/doubling (typical of the short-duration slopes seen here) translates to only 2-dB change in absolute threshold for the same 16-fold increase in signal duration; however, this 2-dB change corresponds to 40% of a 5-dB dynamic range. Both dynamic range and slopes of temporal-integration functions are variable in electric hearing (e.g., see Tables II and III). As shown below, threshold improvement due to temporal integration, expressed as a percentage of dynamic range, may also vary considerably across subjects and electrodes.

The present temporal-integration data are replotted in Fig. 4, with thresholds normalized to the threshold for the longest-duration stimulus (64-pulse or 630.4-ms train), and with threshold change (in dB) expressed as a percentage of the dynamic range (in dB) for each electrode. Computations were based on the dynamic range values listed in Table II. Recall that these were obtained with stimuli (500-ms, 125-Hz 205- μ s/phase pulse trains) relatively similar to the 64-pulse temporal-integration stimulus. The heavy, solid line in Fig. 4 represents data that would be expected for a normal-hearing acoustic listener, based upon a 100-dB dynamic range and a temporal-integration slope of 2.5 dB/doubling of signal duration for durations less than 320 ms.

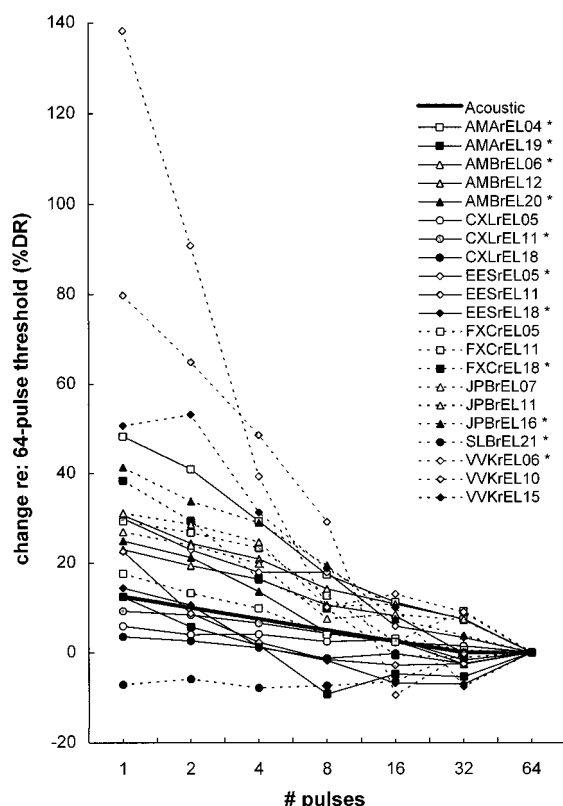


FIG. 4. Temporal-integration functions for 21 electrodes in eight cochlear-implant subjects, shown with thresholds normalized to the threshold for the longest (64 pulse) stimulus and with threshold change expressed as a percentage of the dynamic range for each electrode. Dynamic range values are taken from Table II. Data for individual electrodes are represented by the same symbol-line combinations as in Fig. 2. The heavy, solid line (without symbols) represents data that would be expected for a normal-hearing acoustic listener.

Figure 4 underscores the variability of temporal-integration behavior in electric hearing, both across subjects and for different electrodes in a given subject. Some subjects and electrodes (e.g., CXLrEL18) exhibit less temporal integration (expressed in percent dynamic range) than would be expected in acoustic hearing; however, many others exhibit considerably greater integration than that observed in the acoustic case. Electrode VVKrEL06 again demonstrates the greatest temporal integration, with a maximum threshold shift (approximately 6 dB, see Fig. 2) that is almost 40% greater than the corresponding dynamic range for a 500-ms stimulus (4.5 dB, see Table II). Whereas VVKrEL06 appears to possess unusual characteristics, many other electrodes demonstrate threshold shifts equivalent to 20% or more of dynamic range as stimulus duration decreases from 64 to 4 or fewer pulses. In comparison, an acoustic listener would exhibit threshold shifts corresponding to 10%–15% of dynamic range, depending upon stimulus parameters and the measured value of dynamic range.

The data in Fig. 4 have potentially important implications with respect to cochlear-implant subjects' detection of short-duration speech cues. Many speech processors encode the intensity level of an acoustic stimulus by varying current amplitude, and the input-output function used to map acoustic intensity level to current amplitude for a particular elec-

trode is determined by measuring threshold and comfortable loudness levels for a relatively long-duration (e.g., 500-ms) stimulus. Figure 4 indicates that, for the stimulus parameters used here, the current amplitude necessary to reach threshold for a brief (<40 ms) electric stimulus may correspond to the current at any level from 0% to 50% (or more) of an electrode's dynamic range for a long-duration stimulus. In the extreme case of electrode VVKrEL06, the current amplitude at threshold for a one-pulse stimulus is higher than the current amplitude at MAL for a long-duration stimulus; thus, the use of an input-output function based upon threshold and dynamic range measures for a long-duration stimulus would ensure that a single-pulse stimulus would never be heard.

Consonant discrimination depends strongly on the listener's ability to distinguish differences in the amplitude or spectral characteristics of brief (e.g., <50 ms) acoustic cues embedded in the ongoing speech waveform. Speech processors that use low stimulation rates (e.g., 250 Hz or less in the Nucleus SPEAK and MPEAK processors) encode such cues with relatively few pulses per stimulated electrode. This may result in degraded consonant discrimination in those listeners (like VVK) who have one or more electrodes with steep temporal-integration functions and narrow dynamic ranges. Relatively little is known about the effects of stimulus parameters such as pulse rate, pulse duration, and electrode configuration on temporal-integration behaviors for electric stimuli; thus, it is difficult to predict whether consonant discrimination would be enhanced or degraded by particular speech-processing strategies. Effects of stimulus parameters on temporal-integration slope and dynamic range are likely to vary across individuals; thus, temporal integration may be one of several factors responsible for differences in speech-recognition ability across subjects and for different speech-processing strategies.

Dynamic range as a function of stimulus duration. Our consideration of temporal integration in terms of the dynamic range of individual electrodes led us to question whether dynamic range, itself, is influenced by stimulus duration. To address this question, we evaluated maximum acceptable loudness (MAL) levels for three electrodes in each of four subjects (AMB, EES, FXC, and JPB) using pulse-train stimuli that were identical to those used in constructing temporal-integration functions. MAL estimates were obtained using the ascending method of limits described previously, with stimuli presented continuously at a rate of 1/s. Dynamic range was computed by subtracting the mean 3IFC threshold for each stimulus (data shown in Figs. 2 and 3) from the corresponding, measured MAL. Resulting data are shown in Fig. 5.

Each panel in Fig. 5 displays data for a single electrode: Mean 3IFC-thresholds and MAL estimates are shown as shaded circles and shaded triangles, respectively, referred to the left-hand axis; dynamic range values are indicated by solid diamonds, referred to the right-hand axis. It is evident that the effects of train duration on MAL and dynamic range varied across subjects and electrodes, for the 12 electrodes tested. In general, MAL decreased or remained relatively constant as train duration increased. However, the relative slopes of threshold-duration functions and MAL-duration

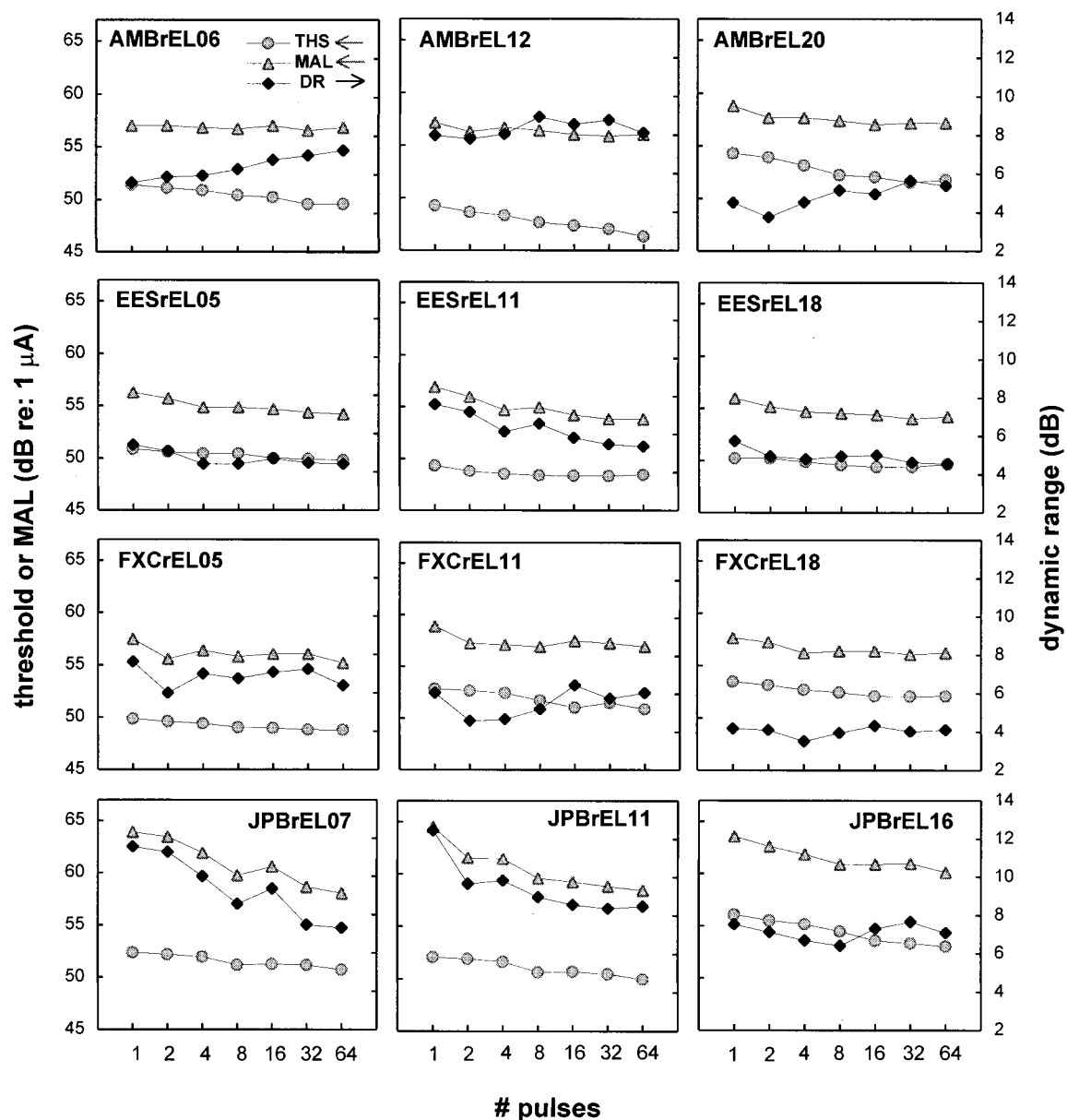


FIG. 5. Threshold, maximum acceptable loudness (MAL), and dynamic range measures as a function of stimulus duration for three electrodes each in four subjects. Each panel shows data for a different electrode, as indicated. Thresholds, represented by shaded circles, are mean 3IFC data replotted from Figs. 2 and 3; MALs, represented by shaded triangles, were obtained using an ascending method-of-limits procedure described in the text. Threshold and MAL data are referred to the left-hand axis. Dynamic range measures computed from threshold and MAL values are shown as solid diamonds, referred to the right-hand axis.

functions varied, with MAL-duration functions being steeper (e.g., JPBrEL07), shallower (e.g., AMBrEL12) or similar in slope (e.g., FXCrEL18) to the threshold-duration functions. This produced dynamic ranges that decreased, increased, or remained relatively constant as the stimulus train was lengthened. In short, stimulus duration had an inconsistent effect on dynamic range for the stimulus parameters used here.

B. Psychometric functions

Figure 6 shows fitted psychometric function data for 11 test electrodes, obtained with the 1-, 2-, 4-, 8-, and 64-pulse stimuli. The data are linear in $\log d'$ versus dB coordinates, as plotted. The 64-pulse data were not used in predictions of

temporal-integration slopes according to the multiple-looks hypothesis (see below), but are shown for completeness.⁸ Details of Fig. 6 are discussed below.

Certainty of slope estimates. Table IV summarizes slope estimates and related statistics for the same psychometric functions. Slopes were computed for 54 of 55 data sets: In one instance (SLBrEL21, one-pulse stimulus), a valid slope estimate could not be obtained because percent-correct responses met the criteria for inclusion in the fitted function at only two levels of the signal. This occurred even though signal levels were sampled at the smallest possible current increments, i.e., 1-CSU steps. Considerable uncertainty was associated with psychometric-function slope estimates for the individually fit (1-, 2-, 4-, 8-, or 64-pulse) functions, as

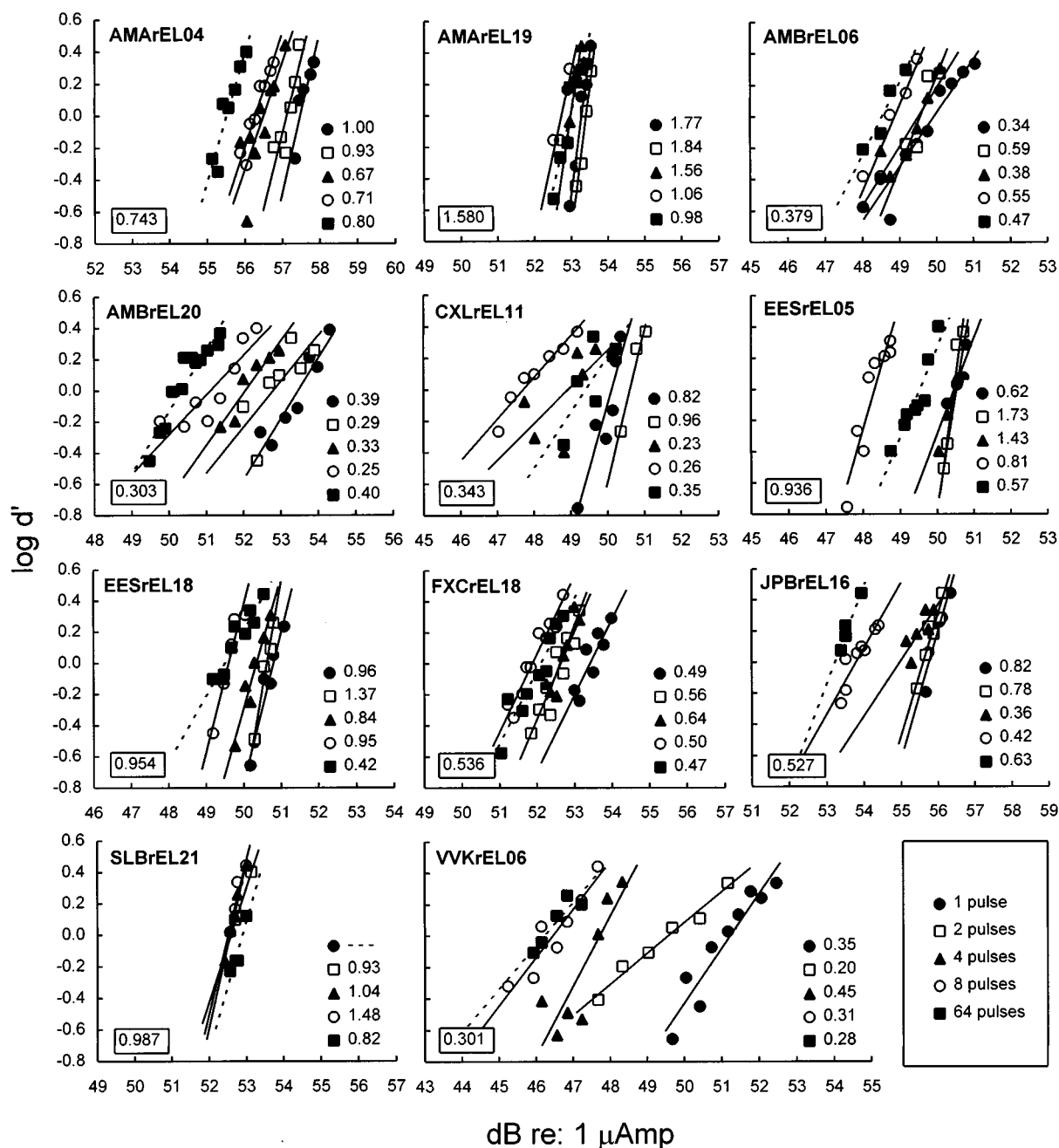


FIG. 6. Psychometric functions for detection for 11 electrodes from eight cochlear-implant subjects. Each panel represents data for a single subject and electrode, as indicated in the panel's top-left corner. Psychometric-function data for 1-, 2-, 4-, and 8-pulse stimuli are shown as filled circles, open squares, filled triangles, and open circles, respectively. Solid lines represent linear regression fits to these short-duration data. Data points and fitted functions for the 64-pulse stimulus are shown as filled squares and dashed lines, respectively. Slope estimates for individual psychometric functions are listed in the lower-right corner of each panel, and the composite slope estimate obtained by simultaneously fitting the 1-, 2-, 4-, and 8-pulse data (see text) is given in the box in each panel's lower-left corner. (Composite functions are not plotted). The scaling of x and y axes has been held constant across panels to emphasize differences in psychometric-function slopes, although the origin varies with electrode.

reflected by their large standard error (se) values and large coefficients of variation (CVs). Coefficients of variation averaged 22% for the 1-, 2-, 4-, and 8-pulse conditions and 19% for the 64-pulse condition. Moreover, large standard errors coupled with the small number of data points comprising individual functions often resulted in unacceptably large confidence intervals (90% CI limits). The certainty of composite slope estimates [comp (1–8)] was considerably better: For these functions, the mean CV was 12%, and confidence

intervals were much narrower. Recall that composite slopes were computed by simultaneously fitting the 1-, 2-, 4-, and 8-pulse data sets with four functions having identical slopes and variable y intercepts (see Sec. I). Composite slope estimates were generally similar to the mean slope estimates for the four short-duration conditions [mean (1–8)]; however, given that underlying psychometric-function slopes are constant with signal duration, the composite estimates likely provide a more accurate estimate of the “true” slope for a

TABLE IV. Individual and composite psychometric-function slopes obtained for 11 electrodes in eight cochlear-implant users. Individual functions for the 1-, 2-, 4-, 8-, and 64-pulse stimuli were estimated by least-squares linear regression fits of the data for a particular stimulus to the equation $\log d' = \alpha'x + b$, where d' is detection performance, x is stimulus amplitude (dB *re*: 1 μ A), and b is a sensitivity constant. Composite slope estimates [comp (1–8)] were obtained by simultaneously fitting four functions $\log d' = \alpha'x + b_i$, with a common slope α' and variable sensitivity constants b_i , to the 1-, 2-, 4-, and 8-pulse data for each electrode. Standard errors of slope estimates [$se(\alpha')$], coefficients of variation [CV(%)], and 90% confidence intervals (90% CI) are listed for each function. For each electrode, the mean of individual short-duration slopes (1-, 2-, 4-, and 8-pulses) is also given [mean (1–8)].

| Electrode | No. pulses | α' | $se(\alpha')$ | CV(%) | 90% CI | Electrode | No. pulses | α' | $se(\alpha')$ | CV(%) | 90% CI |
|-----------|------------|-----------|---------------|-------|------------|-----------|------------|-----------|---------------|-------|------------|
| AMArEL04 | 1 | 1.00 | 0.28 | 28 | 0.35–1.66 | EESrEL18 | 1 | 0.96 | 0.12 | 12 | 0.71–1.21 |
| | 2 | 0.93 | 0.24 | 26 | 0.41–1.45 | | 2 | 1.37 | 0.18 | 13 | 0.86–1.89 |
| | 4 | 0.67 | 0.17 | 25 | 0.35–0.99 | | 4 | 0.84 | 0.11 | 13 | 0.61–1.07 |
| | 8 | 0.72 | 0.09 | 13 | 0.54–0.89 | | 8 | 0.95 | 0.16 | 17 | 0.56–1.33 |
| | 64 | 0.80 | 0.12 | 15 | 0.55–1.05 | | 64 | 0.42 | 0.06 | 13 | 0.32–0.53 |
| | comp (1–8) | 0.74 | 0.09 | 12 | 0.59–0.89 | | comp (1–8) | 0.95 | 0.07 | 7 | 0.83–1.08 |
| AMArEL19 | mean (1–8) | 0.83 | | | | FXCrEL18 | mean (1–8) | 1.03 | | | |
| | 1 | 1.77 | 0.17 | 10 | 1.37–2.17 | | 1 | 0.49 | 0.37 | 31 | 0.32–2.06 |
| | 2 | 1.84 | 0.24 | 13 | 1.14–2.54 | | 2 | 0.56 | 0.09 | 15 | 0.40–0.73 |
| | 4 | 1.56 | 0.10 | 6 | 1.28–1.84 | | 4 | 0.64 | 0.12 | 19 | 0.39–0.89 |
| | 8 | 1.06 | 0.32 | 30 | 0.14–1.98 | | 8 | 0.50 | 0.05 | 8 | 0.41–0.59 |
| | 64 | 0.98 | 0.15 | 15 | 0.68–1.29 | | 64 | 0.47 | 0.06 | 13 | 0.35–0.59 |
| AMBrEL06 | comp 1–8 | 1.58 | 0.13 | 8 | 1.35–1.81 | JPBrEL16 | comp (1–8) | 0.54 | 0.04 | 8 | 0.47–0.60 |
| | mean (1–8) | 1.56 | | | | | mean (1–8) | 0.53 | | | |
| | 1 | 0.34 | 0.04 | 11 | 0.28–0.41 | | 1 | 0.82 | 0.15 | 19 | 0.49–1.14 |
| | 2 | 0.59 | 0.22 | 38 | –0.06–1.23 | | 2 | 0.78 | 0.13 | 17 | 0.51–1.06 |
| | 4 | 0.38 | 0.09 | 24 | 0.18–0.57 | | 4 | 0.36 | 0.14 | 38 | 0.07–0.65 |
| | 8 | 0.55 | 0.11 | 20 | 0.30–0.80 | | 8 | 0.42 | 0.08 | 18 | 0.27–0.57 |
| AMBrEL20 | 64 | 0.47 | 0.10 | 21 | 0.19–0.75 | SLBrEL21 | 64 | 0.63 | 0.10 | 16 | 0.33–0.92 |
| | comp (1–8) | 0.38 | 0.04 | 9 | 0.32–0.44 | | comp (1–8) | 0.53 | 0.07 | 13 | 0.40–0.62 |
| | mean (1–8) | 0.46 | | | | | mean (1–8) | 0.60 | | | |
| | 1 | 0.39 | 0.06 | 16 | 0.26–0.51 | | 1 | a | a | a | a |
| | 2 | 0.29 | 0.11 | 39 | 0.06–0.52 | | 2 | 1.07 | 0.14 | 13 | 0.45–1.00 |
| | 4 | 0.33 | 0.06 | 18 | 0.21–0.46 | | 4 | 1.04 | 0.08 | 8 | 0.84–1.23 |
| CXLEL11 | 8 | 0.25 | 0.05 | 21 | 0.15–0.36 | VVKrEL06 | 8 | 1.48 | 0.41 | 28 | –1.13–4.09 |
| | 64 | 0.40 | 0.05 | 11 | 0.32–0.48 | | 64 | 0.82 | 0.18 | 21 | 0.31–1.33 |
| | comp (1–8) | 0.30 | 0.04 | 12 | 0.24–0.37 | | comp (1–8) | 0.99 | 0.09 | 9 | 0.81–1.16 |
| | mean (1–8) | 0.32 | | | | | mean (1–8) | 1.20 | | | |
| | 1 | 0.82 | 0.16 | 20 | 0.47–1.17 | | 1 | 0.35 | 0.05 | 13 | 0.27–0.43 |
| | 2 | 0.96 | 0.20 | 21 | –0.28–2.20 | | 2 | 0.20 | 0.02 | 8 | 0.16–0.23 |
| EESrEL05 | 4 | 0.23 | 0.14 | 60 | –0.06–0.53 | | 4 | 0.45 | 0.11 | 25 | 0.22–0.67 |
| | 8 | 0.26 | 0.03 | 11 | 0.20–0.33 | | 8 | 0.31 | 0.05 | 17 | 0.20–0.41 |
| | 64 | 0.35 | 0.15 | 43 | 0.03–0.67 | | 64 | 0.28 | 0.07 | 25 | 0.11–0.44 |
| | comp (1–8) | 0.34 | 0.07 | 22 | 0.21–0.47 | | comp (1–8) | 0.30 | 0.03 | 11 | 0.25–0.36 |
| | mean (1–8) | 0.57 | | | | | mean (1–8) | 0.32 | | | |
| | 1 | 0.62 | 0.21 | 34 | 0.01–1.23 | | | | | | |
| | 2 | 1.73 | 0.28 | 16 | 0.93–2.54 | | | | | | |
| | 4 | 1.43 | 1.26 | 88 | 2.24–5.10 | | | | | | |
| | 8 | 0.81 | 0.13 | 16 | 0.55–1.07 | | | | | | |
| | 64 | 0.57 | 0.09 | 15 | 0.40–0.74 | | | | | | |
| | comp (1–8) | 0.94 | 0.17 | 18 | 0.63–1.24 | | | | | | |
| | mean (1–8) | 1.15 | | | | | | | | | |

^aA valid fit could not be obtained because only two data points met the criterion for inclusion.

given electrode. The assumption of invariant slopes is considered below.

Variability of slopes across stimulus conditions and electrodes. Figure 6 illustrates several important aspects of psychometric-function slopes. (Here, we focus on the slopes for 1- to 8-pulse stimuli; slopes of 64-pulse functions are considered briefly below.) First, it can be seen that slopes for the short-duration stimuli varied markedly across electrodes and subjects. Slope estimates for the individually fit 1-, 2-, 4-, and 8-pulse psychometric functions ranged from 0.20 to 1.84 $\log d'/\text{dB } \mu\text{A}$, with a mean (s.d.) value of 0.77 (0.45) $\log d'/\text{dB } \mu\text{A}$. Composite slope estimates computed from the same 1- to 8-pulse data ranged from 0.30 to 1.58 with a mean (s.d.) value of 0.69 (0.40) $\log d'/\text{dB } \mu\text{A}$.

These individual and composite slopes are 2–20 times steeper than psychometric-function slopes reported for detection of tones in quiet by normal-hearing and cochlear-impaired acoustic listeners, which range from 0.08 to 0.15 $\log d'/\text{dB SPL}$ (Carlyon *et al.*, 1990; Viemeister and Wakefield, 1991).

A second aspect of the data shown in Fig. 6 involves differences in slope estimates across the four short-duration conditions. For most electrodes, psychometric function slopes were relatively constant across these stimuli, varying by 33% or less from their mean value. Electrode AMBrEL20, for example, yielded slopes of 0.39, 0.29, 0.33, and 0.25 $\log d'/\text{dB}$ for the 1-, 2-, 4-, and 8-pulse conditions, respectively, and all four individual slope estimates fell

within 22% of their mean value (0.32). Greater variability was exhibited by electrodes EESrEL05 and VVKrEL06. For these electrodes, individual slope estimates differed from the mean short-duration slope by as much as 51% and 40%, respectively; however, slopes did not vary systematically with stimulus duration. Electrodes CXLrEL11 and JPBrEL16 yielded slope estimates that were steeper for the shortest-duration stimuli (single pulses or 2-pulse trains) than for stimuli with slightly longer durations (4- and 8-pulse trains); however, the apparent decrease in psychometric function slope with stimulus duration demonstrated by these electrodes was not statistically significant.⁹

Psychometric-function slopes obtained for the 64-pulse stimulus were generally similar to those obtained for the short-duration stimuli. However, there was one clear exception: For electrode EESrEL18, the 64-pulse psychometric-function slope of $0.42 \log d'/\text{dB}$ was considerably shallower than the slopes estimated for the four short-duration stimuli, which averaged $1.03 \log d'/\text{dB}$. Current levels spanned by psychometric functions (and mean 3IFC thresholds, see Fig. 2) were similar across these conditions; thus, level-dependent changes in neural-response characteristics cannot account for this finding. We can offer no other explanation.

Perhaps the most important aspect of the psychometric functions shown in Fig. 6 is the variation of composite slope estimates across electrodes. As noted above, composite slopes ranged from 0.30 to $1.58 \log d'/\text{dB}$ and were considerably steeper than those reported in acoustic hearing. Composite slopes were similar for pairs of test electrodes in two subjects: AMB's electrodes rEL06 and rEL20 yielded slopes of 0.38 and $0.30 \log d'/\text{dB}$, and EES's electrodes rEL05 and rEL18 yielded slopes of 0.94 and $0.95 \log d'/\text{dB}$. In contrast, composite slope estimates for AMA's electrodes rEL04 and rEL19 differed considerably, with respective values of 0.74 and $1.58 \log d'/\text{dB}$. As mentioned earlier, differences in neural survival may underlie the observed differences in psychometric function slopes across subjects and electrodes. Specific factors that may mediate these survival-related differences are considered below (Sec. III C).

Correlations of slopes with absolute threshold and dynamic range. Sensitive absolute thresholds and wide dynamic ranges have been associated with increased survival of spiral ganglion cells and myelinated peripheral processes, especially when threshold and dynamic range measures are obtained with long phase-duration stimuli (Pfungst *et al.*, 1981; Pfingst and Sutton, 1983; Pfingst *et al.*, 1985; Kawano *et al.*, 1994). Thus, if differences in neural survival are responsible for differences in psychometric-function slopes across subjects and electrodes, then it might be expected that the psychometric-function slopes would vary systematically with one or both of these measures.

To evaluate this possibility, we computed correlations between psychometric-function slopes and two different measures of absolute threshold and dynamic range. Threshold and dynamic range data for 205- $\mu\text{s}/\text{phase}$ (125 Hz, 500 ms) pulse trains, listed in Table II, were available for all 11 electrodes.¹⁰ In addition, threshold and dynamic range data for a longer phase-duration stimulus (1260 $\mu\text{s}/\text{phase}$, 125 Hz, 300 ms), measured as part of another experiment, were

available for six electrodes (AMBrEL06, AMBrEL20, EESrEL05, EESrEL18, FXCrEL18, and JPBrEL16). As noted, above, neural survival may be better predicted by threshold and dynamic range measures at long phase durations; thus, correlations were performed for this subset of the data as well. Scatterplots of composite psychometric-function slopes (from Table IV) versus threshold and dynamic range measures for short and long phase-duration stimuli are shown in Fig. 7. Heavy line segments represent least-squares linear regression fits to the data; corresponding regression parameters and correlation coefficients are indicated in each panel.

Psychometric-function slopes tended to increase with absolute threshold for both the 205- and 1260- $\mu\text{s}/\text{phase}$ conditions, although correlations were relatively weak ($r = 0.479$ and $r = 0.680$, respectively) and failed to reach statistical significance ($p \sim 0.13$ in each case). Dynamic range was more strongly predictive of psychometric-function slope: A moderate negative correlation was obtained for the 205- $\mu\text{s}/\text{phase}$ condition ($r = -0.65$, $\text{df} = 8$, $p = 0.04$), and a strong negative correlation ($r = -0.89$, $\text{df} = 4$, $p < 0.05$) was obtained for the 1260- $\mu\text{s}/\text{phase}$ condition. Although these correlations were based upon a small number of electrodes, they nonetheless suggest a link between slopes of psychometric functions and neural survival.

To summarize, the data shown in Fig. 7 indicate that psychometric-function slopes tend to increase with absolute threshold and are inversely related to dynamic range. This supports our suggestion that the observed variability of psychometric-function slopes across subjects and electrodes may be primarily attributable to differences in auditory-nerve survival.

C. Evaluation of the multiple-looks hypothesis

As discussed earlier, the simple version of the multiple-looks hypothesis predicts an inverse relation between psychometric-function slopes and slopes of the short-duration segments of temporal-integration functions. In terms of the present data, the predicted relation is described by the hyperbolic function

$$10 \log \left(\frac{I_{2T}}{I_T} \right) = \frac{-0.1505}{\alpha'}, \quad (1)$$

where $10 \log(I_{2T}/I_T)$ is the slope of the temporal-integration function in units of dB threshold improvement per doubling of stimulus pulses and α' is the slope of the psychometric function in $\log d'$ versus dB coordinates. The derivation of Eq. (1) is given in the Appendix.

Figure 8 shows a scatterplot of psychometric-function slope versus temporal-integration slope for the 11 electrodes for which both measures were obtained. Data are plotted on log-log axes, resulting in linearization of the hyperbolic function predicted by the multiple-looks hypothesis (solid line). A point representing mean data reported by Carlyon *et al.* (1990) for four normal-hearing acoustic listeners is also shown. Carlyon *et al.* estimated temporal-integration slopes by measuring subjects' thresholds for a single 5-ms, 4-kHz tone pulse and for trains of ten such pulses; they also measured psychometric-function slopes for the same stimuli. The

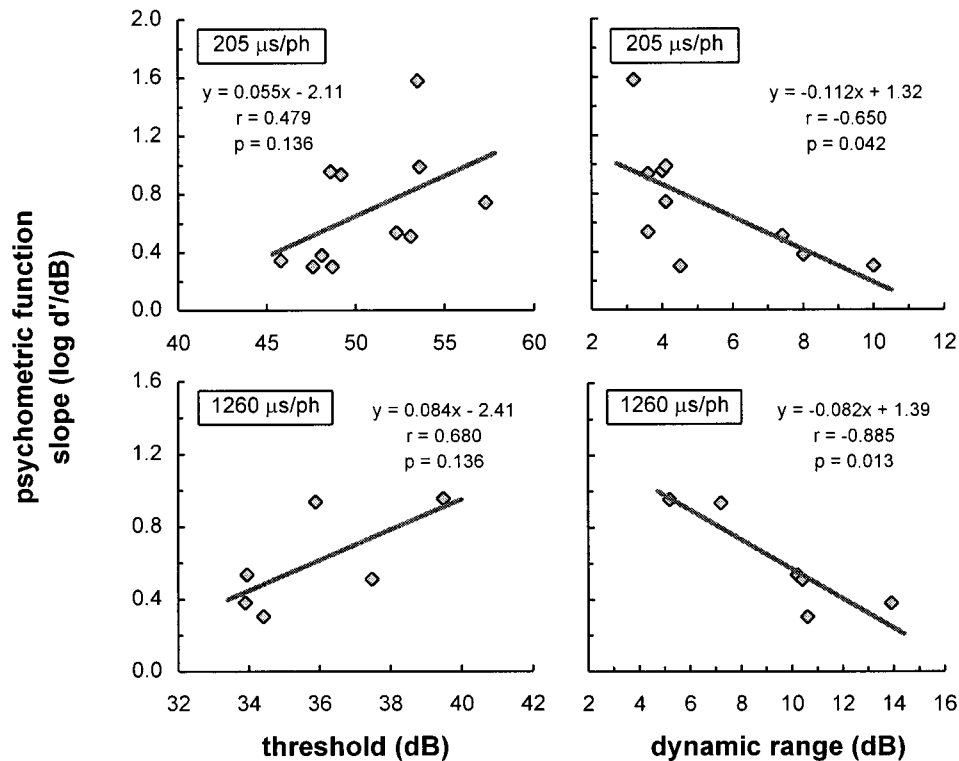


FIG. 7. Scatterplots of composite psychometric-function slope for short-duration (1-, 2-, 4-, and 8-pulse) stimuli versus absolute threshold and dynamic range. Threshold and dynamic range values in the upper panels are for stimuli with short pulse durations (205 $\mu\text{s/ph}$). Threshold data represent 11 electrodes in eight subjects; dynamic range data represent ten electrodes in seven subjects. Corresponding data in the lower panels are for stimuli with long pulse durations (1260 $\mu\text{s/ph}$), and represent six electrodes in four subjects. Heavy lines in each panel represent least-squares linear regression fits to the plotted data. Regression parameters and corresponding correlation coefficients are also given.

data point shown here, which is most comparable to the present data, is for pulse trains with 5-ms interpulse gaps.

The present cochlear-implant data conform fairly closely to the line predicted by the simple version of the multiple-looks model, and the prediction line accounts for 68% of the total variability in these data.^{11,12} The data point representing acoustic data from Carlyon *et al.* (1990) also falls close to the multiple-looks prediction line. There is clearly some error in the multiple-looks' prediction of the present data; however, the model does a reasonable job of describing the observed relation between psychometric-function slope and temporal-integration slope over a relatively wide range of slope values.

III. GENERAL DISCUSSION

A. Comparisons with other published data

Temporal-integration functions obtained from cochlear-implant listeners have been reported in several previous studies. Shannon (1983, 1986) reported thresholds as a function of burst duration for 1000-Hz sinusoids, for one or more electrodes in each of five subjects. Slopes of temporal-integration functions were approximately 2.5 dB/doubling of stimulus duration for three subjects; one function reported for a fourth subject had a slope of approximately 1.5 dB/doubling. Five functions were reported for a fifth subject, and these demonstrated two distinct patterns: Three functions were nearly flat (slopes ~ 0.4 dB/doubling); the other two

showed considerably steeper slopes, but only at very short durations. More recently, Shannon (1990) reported approximately 20 temporal-integration functions from six implanted subjects. Stimulus conditions were not specified, but were presumably the same for all functions. Function slopes varied considerably, from approximately 2.5 dB/doubling to nearly flat. It is difficult to directly compare temporal-integration slopes obtained by Shannon (1983, 1986) with the present data, since stimuli used in the two studies differed in waveshape (sinusoidal versus pulsatile), frequency (1000 Hz versus 100 Hz) and pulse duration (~ 500 μs vs 200 μs). These parameters have all been shown to affect temporal-integration slopes (White, 1984; Pfingst *et al.*, 1991; Moon *et al.*, 1993; Pfingst and Morris, 1993). The data reported by Shannon (1990) suggest the same general degree of intersubject variability in temporal-integration slopes as that observed in the present study.

Moon *et al.* (1993) reported temporal-integration functions that are more directly comparable to the present data. They tested five subjects with the Nucleus implant, using 100-Hz pulse trains comprised of 1–30 biphasic pulses. Slopes of temporal-integration functions became steeper with increasing pulse duration: 96- $\mu\text{s/ph}$ pulse trains produced temporal-integration functions with a mean slope of 0.22 dB/doubling, whereas 1536- $\mu\text{s/ph}$ pulse trains yielded functions with a mean slope of 0.81 dB/doubling. These values are roughly consistent with the mean temporal-

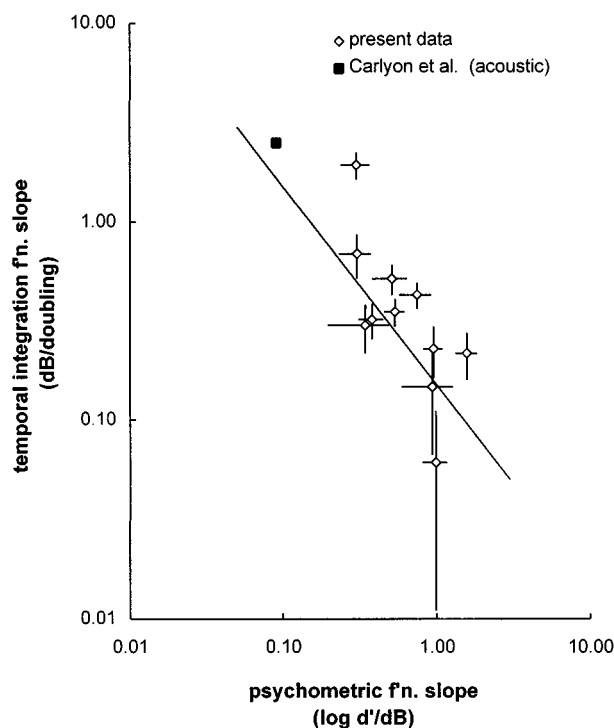


FIG. 8. Scatterplot of temporal-integration slope versus psychometric-function slope for 11 electrodes in eight cochlear-implant subjects for which both measures were obtained. Psychometric-function slopes ($\log d'/\text{dB}$) are the composite short-duration slopes obtained by simultaneously fitting the 1-, 2-, 4-, and 8-pulse data for each electrode with four linear functions having a common slope and variable y intercepts (comp 1–8 slopes in Table IV). Temporal-integration slopes (dB threshold improvement per doubling of number pulses) are those listed in Table III. Horizontal and vertical error bars represent ± 2 standard errors of slope estimates for psychometric functions and temporal-integration functions, respectively. A point representing mean data reported by Carlyon *et al.* (1990) for normal-hearing acoustic listeners is also shown (see text). The solid line represents the equation predicted by the multiple-looks hypothesis [Eq. (1) in the text].

integration slope of 0.42 dB/doubling obtained in the present study for 100-Hz trains of 200- μs /phase biphasic pulses. Note that if the multiple-looks model is correct, Moon *et al.*'s finding that temporal-integration slopes increase with pulse duration predicts that psychometric-function slopes should *decrease* with increasing pulse duration. For example, Eq. (1) predicts that temporal-integration slopes of 0.22 and 0.81 dB/doubling (as reported by Moon *et al.* for 96- and 1536- μs /phase pulses) should correspond to psychometric-function slopes of 0.68 and 0.19 $\log d'/\text{dB}$, respectively. Moon *et al.* did not measure psychometric functions in their study.

Pfingst *et al.* (1991) have similarly observed that temporal-integration functions become steeper with increasing phase-duration of pulse-train stimuli (500- μs /phase versus 5-ms/phase). Although psychometric functions were also obtained, these investigators were unable to detect a systematic effect of phase duration on psychometric-function slopes. As mentioned earlier, it appears that their analyses were limited to qualitative observations. Given this, and considering that detailed analyses of the present data were necessary to demonstrate a relation between psychometric-

function and temporal-integration slopes, their findings are neither surprising nor inconsistent with the present results.

B. Support for the multiple-looks model

The present data generally support the multiple-looks hypothesis of temporal integration. Specifically, they demonstrate an inverse relation between slopes of psychometric functions for detection and slopes of temporal-integration functions that is approximately described by Eq. (1). Equation (1) also predicts the relation between psychometric-function slopes and temporal-integration slopes in acoustic listeners; thus, it accounts relatively well for both electric and acoustic data over a wide range of psychometric-function and temporal-integration slope values.

Deviations of the present data from Eq. (1) may stem largely from the inherent difficulty of measuring psychometric-function slopes and temporal-integration slopes accurately. Both measures are prone to variability imposed by changes in listeners' judgements over time, fluctuations in absolute hearing sensitivity, and statistical limitations of the psychophysical measurement procedures (e.g., binomial variability). Of course, deviations could also stem from factors that are not accounted for by the simple form of the multiple-looks model, or from violations of its assumptions that individual "looks" at the stimulus are independent, equally detectable, and optimally combined.

Additional tests of the multiple-looks model are possible in implanted listeners. For example, as mentioned in the previous section, Pfingst *et al.*, (1991) and Moon *et al.* (1993) have shown that temporal-integration functions increase with biphasic pulse duration in electric hearing, suggesting that slopes of psychometric functions for detection must decrease with pulse duration. If so, and the model is correct, then changes in psychometric-function slope with pulse duration should quantitatively predict changes in temporal-integration slope according to Eq. (1). Importantly, data for a range of pulse durations could be obtained on the same test electrodes, permitting a more direct test of the model than that provided in the present study.

C. Neural factors underlying variability of psychometric-function slopes

Perhaps the most important finding of the present study is that slopes of psychometric functions for detection vary considerably across implant listeners and for different regions of implanted cochleas. The relations discussed earlier, namely the tendency of psychometric-function slopes to increase with absolute threshold, and the inverse correlation between psychometric-function slopes and dynamic range, indicate that differences in psychometric-function slopes across subjects and electrodes may be mediated by differences in the types and relative numbers of surviving cochlear neurons.

A possible explanation for these findings is that low thresholds and wide dynamic ranges reflect the existence of peripheral processes (so-called *dendrites*) in the vicinity of the stimulating electrodes. Threshold detection is presumably mediated by a relatively small number of neural elements

among those closest to the stimulating electrodes; thus, the population neural response to a threshold stimulus should stem primarily from dendritic processes when these exist. Available physiologic data suggest that electrical stimulation of auditory-nerve fibers at their dendritic processes results in shallower rate-intensity functions than stimulation at their central processes (axons) (van den Honert and Stypulkowski, 1984; Javel, 1990). Thus, relatively shallow rate-intensity (RI) functions should characterize neural responses to threshold-level stimuli in regions of good dendritic survival. Other factors (e.g., the stochastic properties of neural responses) being constant, shallow RI functions would be expected to result in shallow psychometric-function slopes.

A second factor to be considered is the spatial density of neural elements stimulated at threshold. Because degeneration occurs in a peripheral-to-proximal sequence (e.g., Spoendlin, 1975; Leake and Hradek, 1988) and loss of dendritic processes is typically greater than loss of spiral ganglion cell bodies in deafened human ears (e.g., Hinojosa and Marion, 1983; Suzuka and Schuknecht, 1988), neural density may be relatively lower for surviving dendrites than for corresponding axons. If so, then growth of the population neural response at threshold should be more gradual in the case of dendritic survival. Both shallow rate-intensity functions and reduced spatial density of surviving neurons are consistent with the relations we have observed here between psychometric function slope and measures of threshold and dynamic range.

IV. CONCLUSIONS

Slopes of temporal-integration functions vary considerably across cochlear-implant listeners and, in some cases, for different electrodes in a particular cochlea. For the stimuli used here (100-Hz trains of 200- μ s/phase biphasic pulses), temporal-integration slopes in implanted listeners are substantially shallower than the 2.5 dB doubling typically reported for acoustic listeners.

Slopes of psychometric functions for detection also vary widely across implant listeners, and for different electrodes in some cochleas. Psychometric-function slopes may be primarily determined by the characteristics of surviving auditory neurons in stimulated regions of the cochlea.

Temporal-integration slopes vary inversely with psychometric-function slopes. This inverse relation is relatively well-described by a hyperbolic function predicted by a simple version of the multiple-looks model of temporal integration. The multiple-looks model appears to account for a wide range of temporal-integration slopes in cochlear-implant and acoustic listeners on the basis of differences in psychometric-function slopes, and is generally supported by the present data.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC00110 and by the Lions 5M International Hearing Foundation. Preliminary findings were presented at the Eighteenth Midwinter Research Meeting of the Association for Research in Otolaryngology (February, 1995). Computer programs

used for data collection were written by John Van Essen; some were modifications of programs provided by Robert Shannon. Tanya Rezac and Lynne Cole assisted in data collection and analysis, and Dorothee Aeppli provided expertise regarding statistical analyses. Very helpful critiques of earlier versions of this manuscript were contributed by Soren Buus and an anonymous reviewer. We are particularly grateful to the eight cochlear-implant subjects who participated in this work.

APPENDIX: DERIVATION OF EQ. (1)

Equation (1) relates slopes of temporal-integration functions to slopes of psychometric functions for detection, according to assumptions of the multiple-looks model of temporal integration. It is derived as follows:

According to the simple form of the multiple-looks hypothesis, $d'_n = \sqrt{n}d'_1$, where d'_n is the detectability of n , independent, optimally combined looks and d'_1 is the detectability of a single look. Assuming $n = bT$, where T is the duration of the stimulus and b is a constant determined by the duration of a look, it follows that

$$d'_T = \sqrt{bT}d'_1.$$

We assume that the psychometric function for one look can be described as $d'_1 = kI^\alpha$. Thus, $d'_T = \sqrt{bT}kI^\alpha$, where k , b , and α are assumed not to depend on n (or T).

The "threshold" value of I is

$$I_T = mT^{-1/2\alpha},$$

where m is a constant determined by the constants k and b , and by the value of d' used to define threshold. For each doubling of duration,

$$\frac{I_{2T}}{I_T} = 2^{-1/2\alpha}$$

and

$$10 \log \left(\frac{I_{2T}}{I_T} \right) = - \frac{1}{2\alpha} (3.01) = - \frac{1.505}{\alpha}.$$

Note that α is the slope of the psychometric function for detection in $\log d'$ versus $\log I$ coordinates. Psychometric-function slopes (α') for the present data were computed in coordinates of $\log d'$ versus $10 \log I$. Thus, $\alpha = 10\alpha'$, and the equation can be rewritten

$$10 \log \left(\frac{I_{2T}}{I_T} \right) = - \frac{0.1505}{\alpha'}.$$

¹The multiple-looks model is not unique in predicting a relation between the form of the psychometric function and temporal integration. As noted by a reviewer, a scheme in which a nonlinearity precedes a temporal integrator and internal noise *could* predict such a relation. However, the relation between the nonlinearity and the form of the psychometric function is uncertain: The uncertainty arises because properties of the internal noise are unknown. In contrast, the simple version of the multiple-looks model, that is discussed here, makes a direct prediction and requires no assumptions about internal noise. Furthermore, there is experimental evidence that the nonlinearity-integration scheme is not tenable as a general account of temporal integration (Viemeister and Wakefield, 1991).

²Current step units (CSUs) are logarithmic units of current amplitude utilized in the Nucleus 22-electrode device. Decibel values equivalent to 1

CSU vary as a function of stimulus level. They generally ranged from 0.1 to 0.3 dB for the stimulus levels used in the present study.

³Collection of complete sets insured that any effects of learning or of time-varying subject state were distributed more-or-less uniformly across stimulus conditions. Random stimulus order would have been preferable in this respect; however, testers could more reliably choose appropriate starting and maximum-safe-stimulation levels when stimuli were tested in a short-to-long sequence, and subjects generally preferred this stimulus order.

⁴The 3 to 4 data sets (21–28 tracks) initially obtained for a particular temporal-integration function typically included 1 or 2 tracks that required replacement because reversals converged poorly and, thus, did not provide a reliable estimate of threshold. In contrast to this, unusual variability of initial thresholds was only a problem for a few stimuli across all 21 functions. Frequently, there was enough time remaining in a subject's scheduled session to collect an additional, partial set of data (in short-to-long stimulus order) after the first 3 to 4 complete sets had been obtained. In such cases, we typically made use of the available time to collect an additional track for each of the short-duration (1-, 2-, 4-, and 8-pulse) stimuli or for each stimulus that yielded initial thresholds that were slightly more variable than others.

⁵The descending–ascending procedure was convenient because it allowed us to determine appropriate stimulus levels on the basis of an initial descending run. Some subjects demonstrated hysteresis in their performance near the middle of the psychometric function for ascending versus descending runs; thus, it seemed appropriate to alternate between these two conditions. We have subsequently compared psychometric function slopes ($\log d'$ /dB) estimated with the descending–ascending procedure to slopes obtained when stimuli were completely randomized across level, and found no systematic differences.

⁶Since psychometric functions were often steep and the size of level increments was limited by the resolution of the implant receiver–stimulator, percent-correct performance sometimes increased from chance to perfect levels in only 3 to 4 steps. Thus, it was important to include as many data points as could be justified in the psychometric-function fits. Use of the confidence-interval criterion allowed us to include points close to chance or perfect performance if the variability of block-to-block percent-correct scores was small. By the same token, it forced us (appropriately) to exclude points in the 60%–90% correct range (which would typically be included) if the corresponding confidence interval was large. The confidence-interval criterion was first used by Buus and Florentine (1991), who showed that it tends to bias resulting slope estimates slightly toward shallower values. The reader is referred to their paper for further discussion of this point.

⁷Gerken *et al.* (1990) have shown that the short-duration portions of acoustic temporal integration functions are approximately linear when plotted in coordinates of \log duration versus \log amplitude.

⁸Note that the relative position of psychometric functions along the abscissa of each panel in Fig. 6 does not always agree with the corresponding temporal-integration function. For example, in the data for CXLrEL11, the one-pulse psychometric function is positioned to the left of the two-pulse psychometric function, suggesting (incorrectly) that the one-pulse threshold was lower than the two-pulse threshold. These discrepancies reflect changes in absolute sensitivity that occurred between the test sessions during which psychometric functions were collected. To insure that such sensitivity shifts did not contaminate the data, temporal-integration functions that were collected over two sessions were monitored for sensitivity shifts, and psychometric functions were always obtained within a single test session (see Sec. I).

⁹Simultaneous fits to the 1-, 2-, 4-, and 8-pulse data were performed using both model I regression (one slope and four intercepts) and model II regression (four slopes and four intercepts). Model II regression did not account for significantly greater variability than model I regression in either case. This indicates that there were no significant differences among the four short-duration slope-estimates for either electrode.

¹⁰Electrode CXLrEL11 did not reach MAL at the output limits of the receiver–stimulator for the 205- μ s/phase stimulus. Because an accurate estimate of dynamic range could not be obtained, this electrode was excluded from the correlation of psychometric function slope versus dynamic range (upper right panel, Fig. 7).

¹¹Because both temporal-integration slopes and psychometric-function slopes were estimated with error, it was appropriate to compute the percentage of total (x plus y) variability accounted for by the multiple-looks prediction line. This computation was performed on the log-transformed data (as plotted in Fig. 8). To determine the “residual” variability (D^2), the distance from each data point to the prediction line was computed

along a vector whose angle was determined by the relative magnitudes of the x and y errors for that point. (If variability was greater in the y dimension, the vector was more nearly vertical; if variability was greater in the x dimension, the vector was more nearly horizontal.) D^2 was taken to be the sum of the squared distances obtained in this manner. The total variability (D_0^2) was computed by determining a weighted sum-of-squares for the x values and a weighted sum of squares for the y values, and summing these quantities. The proportion of variability explained by the prediction line was taken to be $1 - (D^2/D_0^2)$.

¹²Elimination of the data point for electrode VVKrEL06 increased the percentage of variability accounted for by the multiple-looks prediction line from 68.3% to 75.0%. (This electrode's temporal-integration slope was considerably different than that for other electrodes, and could be viewed as an outlier.) The approximate significance (p value) associated with 68.3% explained variability is 0.15; that associated with 75.0% variability is 0.07. Significance values must be viewed as approximate because a theoretical prediction line rather than the best-fitting regression line was used to account for variability.

Buus, S., and Florentine, M. (1991). “Psychometric functions for level discrimination,” *J. Acoust. Soc. Am.* **90**, 1371–1380.

Carlyon, R. P., Buus, S., and Florentine, M. (1990). “Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners,” *J. Acoust. Soc. Am.* **87**, 260–268.

de Boer, E. (1975). “Auditory time constants: A paradox?” in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, Berlin), pp. 141–158.

Eddington, D. K., Doebelle, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. L. (1978). “Auditory prostheses research with multiple channel intracochlear stimulation in man,” *Ann. Otol.* **87**, 1–39.

Florentine, M., Fastl, H., and Buus, S. (1988). “Temporal integration in normal hearing, cochlear impairment and impairment simulated by masking,” *J. Acoust. Soc. Am.* **84**, 195–203.

Forrest, T. G., and Green, D. M. (1987). “Detection of partially filled gaps in noise and the temporal modulation transfer function,” *J. Acoust. Soc. Am.* **82**, 1933–1943.

Gengel, R. W., and Watson, C. S. (1971). “Temporal integration: I. Clinical implications of a laboratory study. II. Additional data from hearing-impaired subjects,” *J. Speech Hear. Disord.* **36**, 213–244.

Gerken, G. M., Bhat, V. K. H., and Hutchison-Clutter, J. H. (1990). “Auditory temporal integration and the power-function model,” *J. Acoust. Soc. Am.* **88**, 767–778.

Green, D. M. (1960). “Auditory detection of a noise signal,” *J. Acoust. Soc. Am.* **32**, 121–131.

Green, D. M. (1985). “Temporal factors in psychoacoustics,” in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, Berlin), pp. 122–140.

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hinojosa, R., and Marion, M. (1983). “Histopathology of profound sensorineural deafness,” *Ann. (N.Y.) Acad. Sci.* **405**, 459–484.

Javel, E. (1990). “Acoustic and electrical encoding of temporal information,” in *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 247–295.

Jeffress, L. A. (1967). “Stimulus-oriented approach to detection re-examined,” *J. Acoust. Soc. Am.* **41**, 480–488.

Jeffress, L. A. (1968). “Mathematical and electrical models of auditory detection,” *J. Acoust. Soc. Am.* **44**, 187–203.

Kawano, A., Seldon, H. L., Pyman, B., and Clark, G. M. (1994). “Intracochlear factors contributing to psychophysical percepts following cochlear implantation: A case study,” *Ann. Otol. Rhinol. Laryngol. (Suppl.)* **166**, 54–57.

Leake, P. A., and Hradek, G. T. (1988). “Cochlear pathology of long term neomycin induced deafness in cats,” *Hearing Res.* **33**, 11–34.

Levitt, H. (1971). “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Am.* **49**, 467–477.

Moon, A. K., Zwolan, T. A., and Pfingst, B. E. (1993). “Effects of phase duration on detection of electrical stimulation of the human cochlea,” *Hearing Res.* **67**, 166–178.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (1988). “The shape of the ear's temporal window,” *J. Acoust. Soc. Am.* **83**, 1102–1116.

- Pfingst, B. E., De Haan, D. R., and Holloway, L. A. (1991). "Stimulus features affecting psychophysical detection thresholds for electrical stimulation of the cochlea. I: Phase duration and stimulus duration," *J. Acoust. Soc. Am.* **90**, 1857–1866.
- Pfingst, B. E., Glass, I., Spelman, F. A., 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.
- Pfingst, B. E., and Morris, D. (1993). "Stimulus features affecting psychophysical detection thresholds for electrical stimulation of the cochlea. II: Frequency and interpulse interval," *J. Acoust. Soc. Am.* **94**, 1287–1294.
- Pfingst, B. E., and Sutton, D. (1983). "Relation of cochlear implant function to histopathology in monkeys," *Ann. (N.Y.) Acad. Sci.* **405**, 224–239.
- Pfingst, B. E., Sutton, D., Miller, J. M., and Bohne, B. A. (1981). "Relation of psychophysical data to histopathology in monkeys with cochlear implants," *Acta Oto-Laryngol.* **92**, 1–13.
- Plomp, R., and Bouman, M. A. (1959). "Relation between hearing threshold and duration for tone pulses," *J. Acoust. Soc. Am.* **31**, 749–758.
- Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," *Hearing Res.* **11**, 157–189.
- Shannon, R. V. (1986). "Temporal processing in cochlear implants," in *Sensorineural Hearing Loss: Mechanisms, Diagnosis, Treatment*, edited by M. H. Collins, T. J. Glatke, and L. A. Harker (University of Iowa, Iowa City), pp. 349–368.
- Shannon, R. V. (1990). "A model of temporal integration and forward masking for electrical stimulation of the auditory nerve," in *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 187–205.
- Shannon, R. V., Adams, D. D., Ferrel, R. L., Palumbo, R. L., and Grandgenett, M. (1990). "A computer interface for psychophysical and speech research with the Nucleus cochlear implant," *J. Acoust. Soc. Am.* **87**, 905–907.
- Spoendlin, H. (1975). "Retrograde degeneration of the cochlear nerve," *Acta Oto-Laryngol.* **79**, 266–275.
- Suzuka, Y., and Schuknecht, H. F. (1988). "Retrograde cochlear neuronal degeneration in human subjects," *Acta Oto-Laryngol. Suppl.* **450**, 1–20.
- van den Honert, C., and Stypulkowski, P. H. (1984). "Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings," *Hearing Res.* **14**, 225–243.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.
- White, M. W. (1984). "Psychophysical and neurophysiological considerations in the design of a cochlear prosthesis," *Audiol. Ital.* **1**, 77–117.
- Zwislocki, J. J. (1960). "Theory of temporal auditory summation," *J. Acoust. Soc. Am.* **32**, 1046–1060.
- Zwislocki, J. J. (1969). "Temporal summation of loudness: An analysis," *J. Acoust. Soc. Am.* **46**, 431–441.

Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa

Roger W. Chan, Ingo R. Titze,^{a)} and Michael R. Titze

National Center for Voice and Speech, Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

(Received 23 September 1996; accepted for publication 15 January 1997)

This paper reports results of further experimentation on a previously developed physical model of the vocal-fold mucosa [Titze *et al.*, J. Acoust. Soc. Am. **97**, 3080–3084 (1995)]. The effects of vocal-fold thickness, epithelial membrane thickness, and prephonatory glottal geometry on phonation threshold pressure were studied. Phonation threshold pressures in the range of 0.13 to 0.34 kPa were observed for an 11-mm-thick vocal fold with a 70- μ m-thick “epithelial” membrane for different “mucosal” fluid viscosities. Higher threshold pressure was always obtained for thinner vocal folds and thicker membranes. In another set of experiments, lowest offset threshold pressure was obtained for a rectangular or a near-rectangular prephonatory glottis (with a glottal convergence angle within about $\pm 3^\circ$). It ranged from 0.07 to 0.23 kPa for different glottal half-widths between 2.0 and 6.0 mm. The threshold for more convergent or divergent glottal geometries was consistently higher. This finding only partially agrees with previous analytical work which predicts a lowest threshold for a divergent glottis. The discrepancy between theory and data is likely to be associated with flow separation from a divergent glottis. © 1997 Acoustical Society of America. [S0001-4966(97)01305-2]

PACS numbers: 43.70.Aj, 43.70.Gr [AL]

INTRODUCTION

Many empirical findings on vocal-fold oscillation have been obtained through the use of physical models and excised larynges (human or animal). Because it is often difficult to isolate and control the parameters of theoretical interest in a human subject, this *in vitro* approach continues to be viable. A physical model of the larynx, for example, constructed with well-defined geometry and biomechanical features, allows for better control of variables and eliminates some variability inherent in human control. Such a model was built earlier for the purpose of measuring phonation threshold pressure (Titze *et al.*, 1995). Phonation threshold pressure (P_{th}) has been defined as the minimum lung pressure required to produce vocal-fold oscillation (Titze, 1988, 1992). It was found to be consistently higher for oscillation onset than for oscillation offset both experimentally (Baer, 1975; Titze *et al.*, 1995) and analytically (Lucero, 1995). Thus *onset* P_{th} can be defined as the minimum lung pressure that initiates vocal-fold oscillation from rest, while *offset* P_{th} (“minimum sustaining pressure” in Lucero, 1995) can be defined as the minimum lung pressure that sustains vocal-fold oscillation after it has begun.

This study represents further experimentation on the previously developed physical model (Titze *et al.*, 1995), in order to explore some untested aspects of a small-amplitude oscillation theory (Titze, 1988). Specifically, the effects of vocal-fold thickness, epithelial membrane thickness and prephonatory glottal convergence angle on P_{th} were explored.

The apparent trade-off relation between epithelial membrane thickness and mucosal fluid viscosity was also studied.

I. PREVIOUS ANALYTICAL AND EXPERIMENTAL RESULTS

Assuming small-amplitude oscillation conditions and a surface wave propagating on the cover of a body-cover model of the vocal folds, Titze (1988) derived some analytical expressions for phonation threshold pressure. For a rectangular prephonatory glottal geometry, the expression was

$$P_{th} = (k_t/T)Bc\xi_0, \quad (1)$$

where k_t is the transglottal pressure coefficient (about 1.1 as determined by Scherer, 1981; Scherer and Guo, 1991), T is the vertical vocal-fold thickness, B is the viscous damping coefficient of vocal fold tissues, c is the mucosal wave propagation velocity, and ξ_0 is the prephonatory glottal half-width. Equation (1) predicts that P_{th} varies linearly with viscous damping in vocal-fold tissues, mucosal wave velocity, and prephonatory glottal half-width. It also predicts that P_{th} varies inversely with vocal-fold thickness.

Previous experimental results on the physical model showed that P_{th} indeed increases with mucosal fluid viscosity and prephonatory glottal half-width in a roughly linear fashion (Titze *et al.*, 1995), supporting Eq. (1), but linearity was not preserved to zero glottal width. Rather, P_{th} was actually lowest for a small positive value of glottal half-width (between 0.0 and 1.0 mm), and rose when ξ_0 approached zero. This nonlinearity might be the consequence of viscous pressure losses which become dominant as ξ_0 gets small (Lucero, 1996). Another unpredicted result was the aforementioned hysteresis effect, with offset P_{th} being consistently lower than onset P_{th} . In response to this finding, Luc-

^{a)}Wilbur James Gould Voice Research Center, The Denver Center for the Performing Arts, Denver, CO 80204.

ero (1995) extended the analytical approach to include large-amplitude oscillations. He came up with the same expression for onset P_{th} [Eq. (36) in Lucero's analysis reduces to Eq. (1) here when static vocal-fold displacement $\bar{\xi}=0$ for a rectangular prephonatory glottis]. Lucero then showed that offset P_{th} was equal to half the onset P_{th} for a rectangular glottis, as a consequence of nonlinear changes in effective aerodynamic damping on the vocal folds.

Returning to our former experimental results, there was an apparent trade-off between epithelial membrane thickness and mucosal fluid viscosity (Titze *et al.*, 1995). For the physical model with a 200- μm -thick membrane and at a physiological range of fluid viscosity (on the order of 500 cP according to Zhu and Mow, 1990), P_{th} was excessively high in relation to P_{th} values for human phonation reported in the literature (on the order of 0.3 kPa at 100 Hz; e.g., Verdolini-Marston *et al.*, 1990, 1994). Since the experimental membrane thickness was four times that of the epithelial layer in human vocal folds (50 μm according to Hirano, 1975), it was concluded that realistic P_{th} values could be obtained only with thinner membranes. On the other hand, lower fluid viscosity could be used to compensate for thicker membranes.

Now let us turn our attention to the more general case of nonrectangular (convergent and divergent) prephonatory glottal geometries. The analytical expression derived by Titze (1988) was

$$P_{th} = (2k_t/T)(Bc)\xi_{01}^2/(\xi_{01} + \xi_{02}), \quad (2)$$

where ξ_{01} and ξ_{02} are inferior and superior prephonatory glottal half-widths, respectively. Note that Eq. (2) reduces to Eq. (1) for a rectangular glottis, where $\xi_{01} = \xi_{02}$. For a divergent glottis ($\xi_{01} < \xi_{02}$), P_{th} is lower than that for a rectangular glottis, while P_{th} is higher for a convergent glottal geometry ($\xi_{01} > \xi_{02}$) because of the square power of ξ_{01} .

Lucero's (1995) large-amplitude analysis of this more general case yielded the following expression:

$$P_{th} = (k_t/2T)(Bc)(\xi_{01} + \bar{\xi})(1 + \sqrt{1 - a^2}), \quad (3)$$

where $\bar{\xi}$ is the static vocal-fold displacement and a is a normalized oscillation amplitude. Whether threshold for onset or offset is represented by this equation depends on the normalized amplitude of oscillation (e.g., $a=0$ for oscillation onset).

The static displacement $\bar{\xi}$ is also a function of prephonatory glottal geometry and is given by

$$\bar{\xi} = (Bc/2KT)(\xi_{01} - \xi_{02})(1 + \sqrt{1 - a^2})/\sqrt{1 - a^2}, \quad (4)$$

where K is the lumped effective stiffness per unit area of the vocal-fold mucosa. The second term in parentheses reveals that $\bar{\xi}$ is positive for a convergent glottis and negative for a divergent glottis. With this static displacement, Eqs. (3) and (4) predict that P_{th} is lowest for the divergent glottal geometry and highest for the convergent glottis, a result similar to that of Eq. (2). Lucero's (1993) dynamic analysis of the two-mass model of vocal folds (Ishizaka and Flanagan, 1972) also predicted the same result.

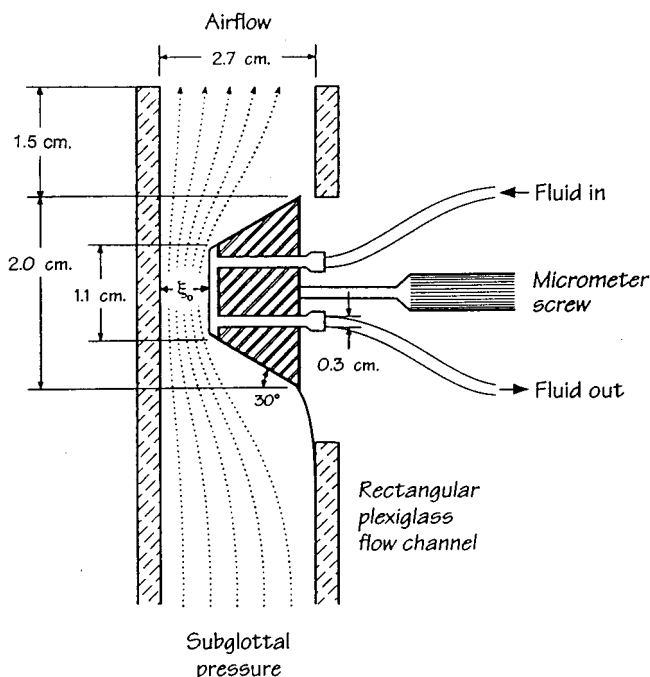


FIG. 1. Sketch of the physical model of the vocal fold and the glottal airway with a rectangular prephonatory glottal geometry. Transverse dimension of the airway is 2.3 cm (after Titze *et al.*, 1995).

II. METHOD

The physical model of the vocal-fold mucosa used in this study was the same one described by Titze *et al.* (1995), where construction details can be found. Briefly, it consisted of a stainless-steel trapezoidal vocal fold "body" and a "mucosa" made of a silicone membrane (the "epithelium") encapsulating a fluid of varying viscosities (the "superficial layer of lamina propria") (Fig. 1). Note that the dimensions of the model are slightly bigger than those of human, since it was designed not to scale to exact human values but to test general relationships. Also note that the relatively short and wide supraglottal "vocal tract" section of the airway (as shown in Fig. 1) provides insignificant acoustic inertive load to the model (an order of magnitude smaller than the inertance of a typical epilarynx or a typical full vocal tract).

Vocal-fold thickness of the model was adjustable by using "bodies" of different sizes. "Epithelial" membrane thickness was varied by controlled dipping of a silicone dispersion fluid and measured by a digital caliper. Thickness of membranes used in the previous study was on the order of 200 μm (as mentioned), but it was possible to manufacture intact membranes as thin as 70 μm in this study. This was much closer to the 50- μm epithelial thickness in human vocal folds (Hirano, 1975).

The vocal-fold position within a rectangular Plexiglas airflow channel was controlled by a micrometer, "adducting" the fold against one wall of the channel, as in a hemilarynx setup (see Fig. 1). Prior to fine adjustment of the prephonatory glottal half-width, the appropriate fluid was injected slowly into the "mucosa" via a syringe. The volume of fluid encapsulated between the "epithelium" and the "body" was controlled by video tracing as in the previous

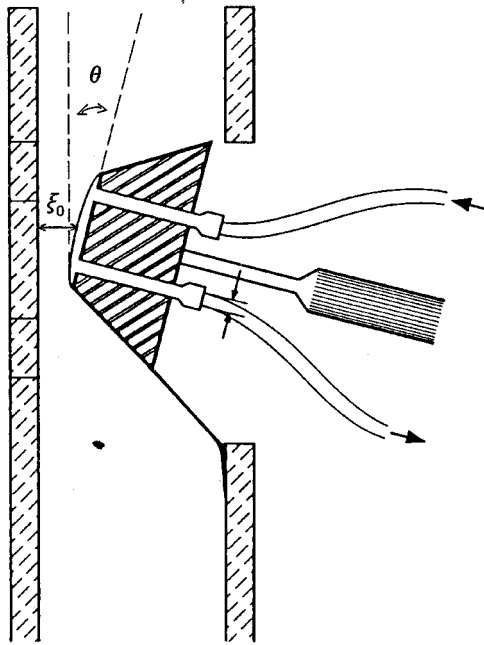


FIG. 2. Schematic of the physical model with a divergent prephonatory glottal geometry. Note that the glottal convergence angle (θ) is negative in a divergent glottis by definition. Also note that the medial “mucosal” surface is actually flat.

study. Glottal geometry was adjusted by tilting the micrometer while maintaining the mid-point prephonatory glottal half-width (ξ_0) constant (Fig. 2). This geometry was expressed as a glottal convergence angle (θ). Note that a flat medial “mucosal” surface was approximated carefully during the process of attaching the silicone membrane to the “body” and also during fluid injection and removal.

The entire model was then mounted onto a pipe that supplied compressed air to the flow channel. Pressure was regulated by a Fairchild model 10 regulator (0–2 psi) and measured with an open-ended water manometer (Dwyer, 60 cm) with a resolution of 0.2 cm H_2O (0.02 kPa). Laminar flow was supplied by this subglottal experimental setup as shown elsewhere (Alipour *et al.*, 1996). Throughout the experiments, pressure was varied by adjusting the regulator as slowly as possible, in order to minimize any inertial effects and to facilitate accurate manometer reading and accurate observation of “mucosal” oscillation (typical pressure rise and fall times were on the order of 1 Pa/s). A video camera was mounted directly above the model to magnify and record the oscillation.

Two sets of experiments were done. In the first set, a rectangular glottis was used with four variations of “epithelial” membrane thickness and “mucosal” fluid viscosity (see Table I). For each variation, onset and offset P_{th} were measured for two different vocal-fold thicknesses (7.5 and 11.0 mm). A glottal half-width of 2.0 mm was maintained throughout the experiments. This $2 \times 2 \times 2$ design allowed the effect of vocal-fold thickness on P_{th} to be studied, as well as the trade-off relation between membrane thickness and fluid viscosity.

In the second set of experiments, prephonatory glottal half-widths of 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, and 6.0 mm were

TABLE I. Four variations of membrane thickness and fluid viscosity used in the first set of experiments. Pure water with a viscosity of 1.0 cP was used for sets A and B, and a 0.5% mixture (by weight) of sodium carboxymethyl cellulose (CMC) powder with H_2O was used for sets C and D.

| | Set A | Set B | Set C | Set D |
|--------------------------------|-------|-------|-------|-------|
| Membrane thickness (μm) | 70 | 210 | 70 | 210 |
| Fluid viscosity (cP) | 1.0 | 1.0 | 471 | 471 |

adopted and prephonatory glottal convergence angle was varied from about -10° to $+10^\circ$ for each of these widths. Other variables were maintained at constant values: vocal-fold thickness at 15 mm, membrane thickness at 130 μm , and tissue viscosity at 1.0 cP (pure water). Note that the membrane thickness was not the thinnest obtainable at the time of experimentation, but was less prone to rupture when angle was varied and repeated trials were taken. Onset and offset P_{th} were measured for each of the above conditions whenever stable oscillation was achieved (it was not always possible to achieve stable oscillation with smaller glottal half-widths and bigger angles because the “mucosa” collided partially with the opposing Plexiglas wall).

III. RESULTS AND DISCUSSION

Figure 3 shows results of the first set of experiments, where onset and offset P_{th} are plotted against vocal-fold thickness with “epithelial” membrane thickness and “mucosal” fluid viscosity as parameters. Each data point represents an averaged measurement of three to four successive trials. As reported in the previous study, experimental error or the overall repeatability of the data was within about ± 0.02 kPa (Titze *et al.*, 1995).

There was again a hysteresis effect as observed before, with offset P_{th} (solid lines) consistently lower than corresponding onset P_{th} (dashed lines). The difference was smallest (about 0.04 kPa) for a thin membrane and a low viscosity and largest (about 0.2 to 0.4 kPa) for a thick membrane and a high viscosity. Interestingly, the ratios between offset and onset P_{th} are all very similar (about 0.75 to 0.80).

In every case, the thicker vocal fold had a lower P_{th} than the thinner one, across all four variations of membrane thickness and fluid viscosity. This finding agrees with the general prediction from Eq. (1). Like the hysteresis effect, the difference was smallest (about 0.03 kPa) for a thin membrane and a low viscosity and largest (about 0.4 to 0.6 kPa) for a thick membrane and a high viscosity. Unlike the hysteresis effect, however, the ratios between P_{th} for the two vocal-fold thicknesses are less constant, ranging from 0.65 to 0.85. This may suggest that the relation between P_{th} and vocal-fold thickness is more than simply inverse, and that previous analytical work might be inadequate in describing its interaction with other parameters.

Figure 3 also shows clearly the trade-off relation between membrane thickness and fluid viscosity. For a non-physiological 210- μm -thick membrane and a physiological fluid viscosity of 471 cP, P_{th} was around 1.0 kPa, considerably above measured P_{th} values for human phonation (about 0.3 kPa at 100 Hz according to the literature, e.g., Verdolini-

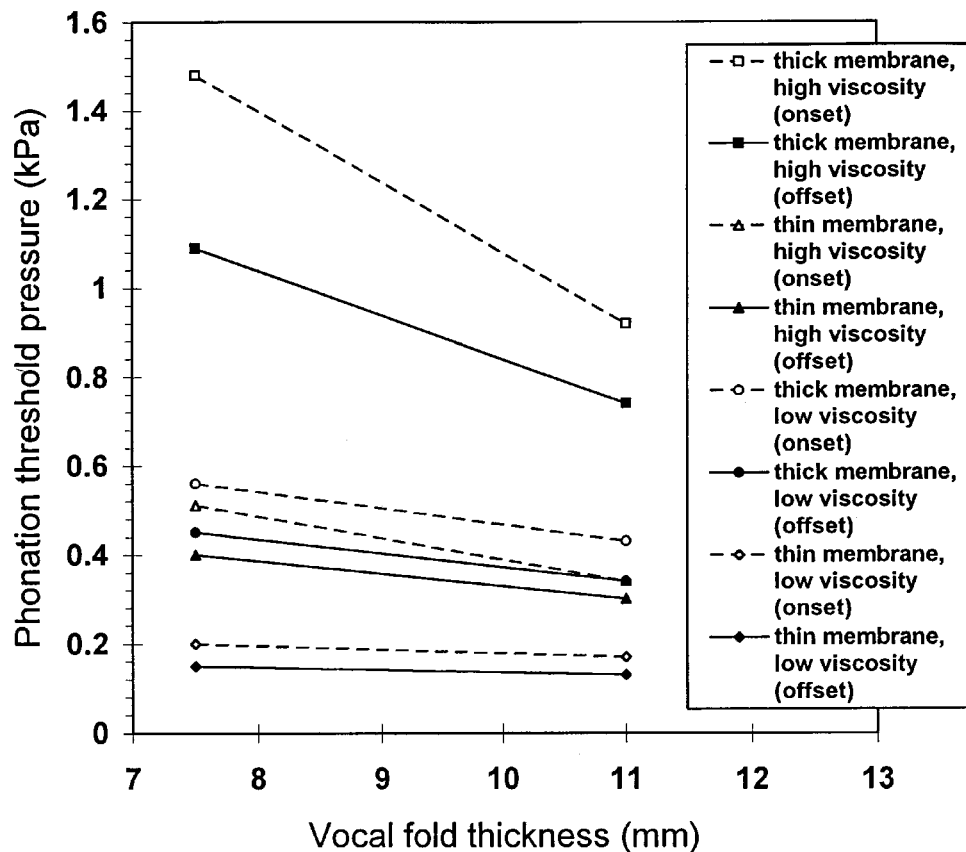


FIG. 3. Phonation threshold pressure as a function of vocal-fold thickness, “epithelial” membrane thickness and “mucosal” fluid viscosity. Each data point represents an average of measurements from three to four successive experimental trials. Prephonatory glottis was rectangular with a half-width of 2.0 mm. “Epithelial” membrane thickness was 70 μm (thin) or 210 μm (thick). “Mucosal” fluid was pure water with a viscosity of 1.0 cP (low) or sodium CMC solution with a viscosity of 471 cP (high).

Marston *et al.*, 1990, 1994). However, when a near-physiological 70- μm -thick membrane was used, realistic P_{th} at about 0.3 to 0.5 kPa were obtained with the physiological fluid viscosity of 471 cP. Comparably realistic P_{th} values were also obtained for a directly opposite and yet totally nonphysiological condition, with a 210- μm -thick membrane and pure water. A practical implication of this finding is that hydration treatments on benign vocal-fold lesions (Verdolini-Marston *et al.*, 1994) might facilitate the ease of phonation by compensating for the unfavorable increase of P_{th} when lesions (e.g., vocal nodules or epithelial hyperplasia) cause an increase in epithelial thickness (and mass) (Colton and Casper, 1996; Gould *et al.*, 1995).

Results of the second set of experiments are shown in Fig. 4, where offset P_{th} is plotted against prephonatory glottal convergence angle, with glottal half-width as the parameter. Only data on offset P_{th} are presented because data on onset P_{th} show a similar pattern. Again, each data point represents an average of measurements from three to four successive trials.

It can be seen that P_{th} generally increases with prephonatory glottal half-width at smaller ξ_0 values (<3.0 mm) (vertical separation of the curves), a finding already discussed in the previous study (Titze *et al.*, 1995). As ξ_0 rises above 3.0 mm, however, the relationship becomes nonlinear where P_{th} remains basically unchanged (within experimental error) for both the rectangular and the slightly divergent

($<5^\circ$) glottis. The pattern becomes more obscure for the convergent glottis. For example, P_{th} for a ξ_0 of 3.5 mm is slightly higher than that for 5.0 and 6.0 mm at a convergence angle of about 2° , but much higher at about 5° . All these nonlinearities are not predicted by Eqs. (1) and (2). Further analytical work is needed and separate expressions are probably required to better describe the relationship in cases with different prephonatory glottal geometries.

P_{th} also changes considerably with prephonatory glottal convergence, the major variable of interest in this set of experiments, except for the case of smallest glottal half-width (2.0 mm) where vocal fold collision occurred at all but the smallest convergence and divergence angles. Note that P_{th} ranges only between 0.07 and 0.08 kPa for this condition and should be regarded as constant, given an estimated experimental error of ± 0.02 kPa.

With experimental error being considered, Fig. 4 shows that the lowest P_{th} was obtained for zero or near-zero glottal convergence angle (within about $\pm 3^\circ$), i.e., a rectangular or near-rectangular prephonatory glottis. P_{th} for convergent or divergent glottal geometries were consistently higher, except for cases where stable oscillation was not achieved.

This finding only partially agrees with Eqs. (2), (3), and (4), which predict that P_{th} is lowest for the divergent glottis and highest for the convergent one. This discrepancy between theory and data cannot be accounted for by the slight difference in minimal constriction width (<1.0 mm) be-

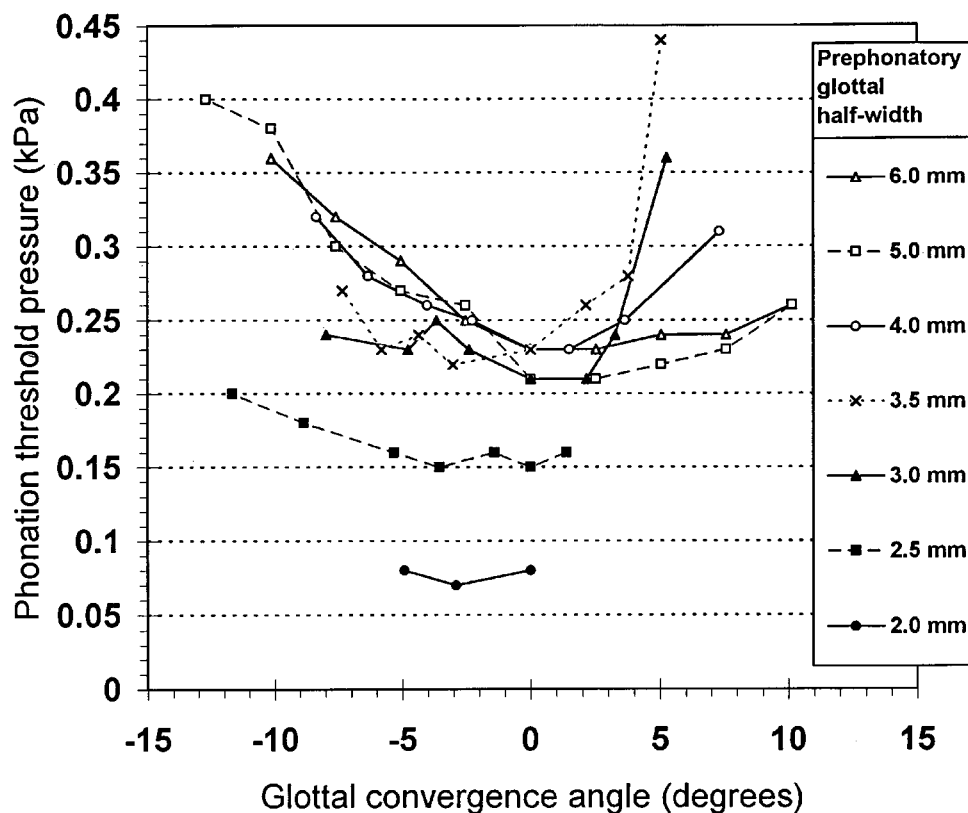


FIG. 4. Offset phonation threshold pressure as a function of prephonatory glottal half-width and glottal convergence angle.

tween a rectangular glottis and a nonrectangular one, because results in Fig. 4 show that there is no significant change in P_{th} associated with such a magnitude of difference in minimal constriction width. On the other hand, the discrepancy between theory and data might be interpreted in terms of flow separation effects in a divergent glottis. Pelorson *et al.* (1994, 1995) described glottal fluid mechanics with a theoretical quasisteady flow model and predicted flow separation for a divergent glottis, in particular a moving flow-separation point during closing phase of the vibratory cycle. Guo and Scherer (1993) used the finite-element method to simulate two-dimensional steady airflow and pressure through the glottis. Flow separation was found to occur at a short distance downstream of the point of minimal constriction in a divergent glottis. They simulated one-dimensional pressure profiles along the glottal airway for different prephonatory glottal geometries (Fig. 8 in Guo and Scherer, 1993). Results showed that more divergent glottal angles are associated with a less negative *overall* pressure along the glottis, even though the pressure at glottal entrance for a 10° divergent glottis is slightly more negative than that for the 5° case. This is presumably because the point of flow separation moves upstream with an increase in glottal divergence, a result implicated in Pelorson *et al.*'s (1994) model of moving flow-separation point. A similar pattern of pressure change was also found in their later experimental study (Pelorson *et al.*, 1995). Hence, a smaller asymmetric driving force is available for the more divergent glottis in each vibratory cycle, given the same subglottal pressure. An increase of P_{th} with glottal divergence is therefore likely in terms of the flow separation effects.

Indeed, it could be possible that the prephonatory glottis diverges to a point that oscillation cannot be established at all, because of the failure to build up sufficient alternate positive and negative pressures or asymmetric driving force necessary for self-sustained oscillation (cf. Scherer, 1981; Titze, 1988). A "critical divergent angle" might therefore exist and should be explored in future studies.

IV. SUMMARY AND CONCLUSION

Experiments were conducted on a previously described physical model of the vocal-fold mucosa. The effects of vocal-fold thickness, epithelial membrane thickness, and prephonatory glottal geometry on phonation threshold pressure were quantified. As predicted by previous analytical result, higher P_{th} was always obtained for a thinner vocal fold, although the relation between vocal-fold thickness and P_{th} might not be exactly inverse.

An increase in "epithelial" membrane thickness was also found to raise P_{th} , which could be compensated by a decrease in "mucosal" fluid viscosity. As far as the prephonatory glottal geometry is concerned, lowest P_{th} was always observed for a rectangular prephonatory glottis, while it was predicted from previous analytical work that a divergent glottis should have the lowest threshold pressure. The discrepancy between theory and data could be interpreted in terms of flow separation effects in a divergent glottis. More analytical work has to be done to further explore this topic in terms of its relationship to phonation threshold pressure.

Clinically, the observed trade-off relation between membrane thickness and fluid viscosity suggests that hydration

treatments on benign vocal-fold lesions (Verdolini-Marston *et al.*, 1994) might facilitate the ease of phonation by compensating for the unfavorable increase of phonation threshold pressure when lesions (e.g., vocal nodules or epithelial hyperplasia) cause an increase in epithelial thickness (and mass) (Colton and Casper, 1996; Gould *et al.*, 1995). Results of the second set of experiments suggest that prephonatory glottal geometry should be considered and carefully monitored in vocal-fold medialization types of phonosurgery (e.g., vocal-fold augmentation with injectable biomaterials, or thyroplasty type I), in order to obtain a nearly rectangular prephonatory glottal geometry. This geometry likely facilitates the ease of phonation. Highly convergent or divergent prephonatory glottal geometries are associated with higher phonation threshold pressures and should be carefully avoided during phonosurgery.

ACKNOWLEDGMENTS

This study was supported by Grant No. P60 DC00976 from the National Institutes on Deafness and Other Communication Disorders. The authors thank the reviewers for their helpful comments.

Alipour, F., Scherer, R., and Knowles, J. (1996). "Velocity distributions in glottal models," *J. Voice* **10**, 50–58.

Baer, T. (1975). "Investigation of phonation using excised larynges," Ph.D. dissertation, MIT, Cambridge, MA.

Colton, R. H., and Casper, J. K. (1996). *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment* (Williams and Wilkins, Baltimore), pp. 58–77.

Gould, W. J., Rubin, J. S., and Yanagisawa, E. (1995). "Benign vocal fold pathology through the eyes of the laryngologist," in *Diagnosis and Treatment of Voice Disorders*, edited by J. S. Rubin, R. T. Sataloff, G. S. Korovin, and W. J. Gould (Igaku-Shoin, New York), pp. 137–151.

Guo, C. G., and Scherer, R. C. (1993). "Finite element simulation of glottal flow and pressure," *J. Acoust. Soc. Am.* **94**, 688–700.

Hirano, M. (1975). "Phonosurgery: Basic and clinical investigations," *Otologia (Fukuoka)* **21**, 239–240.

Ishizaka, K., and Flanagan, J. L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* **51**, 1233–1268.

Lucero, J. C. (1993). "Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region," *J. Acoust. Soc. Am.* **94**, 3104–3111.

Lucero, J. C. (1995). "The minimum lung pressure to sustain vocal fold oscillation," *J. Acoust. Soc. Am.* **98**, 779–784.

Lucero, J. C. (1996). "Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis," *J. Acoust. Soc. Am.* **100**, 2551–2554.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (1994). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," *J. Acoust. Soc. Am.* **96**, 3416–3431.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., and Bailliet, H. (1995). "Description of the flow through in-vitro models of the glottis during phonation," *Acta Acust.* **3**, 191–202.

Scherer, R. C. (1981). "Laryngeal fluid mechanics: Steady flow considerations using static models," Ph.D. dissertation, The University of Iowa, Iowa City, IA.

Scherer, R. C., and Guo, C. G. (1991). "Generalized translaryngeal pressure coefficient for a wide range of laryngeal configurations," in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego), pp. 83–90.

Titze, I. R. (1988). "The physics of small-amplitude oscillation of the vocal folds," *J. Acoust. Soc. Am.* **83**, 1536–1552.

Titze, I. R. (1992). "Phonation threshold pressure: A missing link in glottal aerodynamics," *J. Acoust. Soc. Am.* **91**, 2926–2935.

Titze, I. R., Schmidt, S. S., and Titze, M. R. (1995). "Phonation threshold pressure in a physical model of the vocal fold mucosa," *J. Acoust. Soc. Am.* **97**, 3080–3084.

Verdolini-Marston, K., Titze, I. R., and Druker, D. G. (1990). "Changes in phonation threshold pressure with induced conditions of hydration," *J. Voice* **4**, 142–151.

Verdolini-Marston, K., Sandage, M., and Titze, I. R. (1994). "Effect of hydration treatments on laryngeal nodules and polyps and related voice measures," *J. Voice* **8**, 30–47.

Zhu, W. B., and Mow, V. C. (1990). "Viscometric properties of proteoglycan solutions at physiological concentrations," in *Biomechanics of Diarthrodial Joints*, edited by V. C. Mow, A. Ratcliffe, and S. Woo (Springer-Verlag, New York), pp. 313–344.

Articulatory strengthening at edges of prosodic domains

Cécile Fougeron

Phonetics Laboratory, Department of Linguistics, University of California at Los Angeles, Los Angeles, California 90095-1543 and Institut de Phonétique, Paris Sorbonne–Nouvelle, URA 1027, France

Patricia A. Keating

Phonetics Laboratory, Department of Linguistics, University of California at Los Angeles, Los Angeles, California 90095-1543

(Received 16 April 1996; accepted for publication 25 February 1997)

In this paper it is shown that at the edges of prosodic domains, initial consonant and final vowels have more extreme (less reduced) lingual articulations, which are called articulatory strengthening. Linguopalatal contact for consonants and vowels in different prosodic positions was compared, using reiterant-speech versions of sentences with a variety of phrasings read by three speakers of American English. Four prosodic domains were considered: the phonological word, the phonological (or intermediate) phrase, the intonational phrase, and the utterance. Domain-initial consonants show more linguopalatal contact than domain-medial or domain-final consonants, at three prosodic levels. Most vowels, on the other hand, show less linguopalatal contact in domain-final syllables compared to domain-initial and domain-medial. As a result, the articulatory difference between segments is greater around a prosodic boundary, increasing the articulatory contrast between consonant and vowels, and prosodic domains are marked at both edges. Furthermore, the consonant initial strengthening is generally cumulative, i.e., the higher the prosodic domain, the more linguopalatal contact the consonant has. However, speakers differed in how many and which levels were distinguished in this way. It is suggested that this initial strengthening could provide an alternative account for previously observed supralaryngeal declination of consonants. Acoustic duration of the consonants is also affected by prosodic position, and this lengthening is cumulative like linguopalatal contact, but the two measures are only weakly correlated. © 1997 Acoustical Society of America. [S0001-4966(97)04106-4]

PACS numbers: 43.70.Aj, 43.70.Fq [AL]

INTRODUCTION

It is by now well-established that prosody affects articulation. Beckman and Edwards (1994, p. 8) define prosody as “the organizational framework that measures off chunks of speech into countable constituents of various sizes.” These constituents are called prosodic domains, and their organization is called prosodic structure. Prosodic structure plays an important role in the realization of the “content” of speech sounds (Beckman and Edwards, 1994; Pierrehumbert and Beckman, 1988, p. 116; Fujimura, 1990b, p. 325). Beckman and Edwards distinguish two kinds of locations within prosodic domains that lead to differences in the articulation of content. One location is the head, or most prominent part, of a domain. For example, a nuclear accented syllable is the head of an intermediate phrase, and its vowel can have greater duration, lingual displacement towards the target, or velocity than other vowels (Beckman and Edwards, 1994; de Jong, 1995). The other location is next to the boundaries of the domain, the initial and final edge positions. For example, lengthening, a temporal change, has been found to occur at the initial and final edges of prosodic domains (e.g., Oller, 1973 for English; Byrd *et al.*, 1997 for Tamil). Our interest here is to add to the literature on domain edges, particularly the relatively small literature concerning spatial changes at edges.

For word edges, several studies have shown that articulations of the tongue, lips, velum, and glottis differ in mag-

nitude in word-initial versus non-initial position (e.g., Fromkin, 1965; Vaissière, 1988; Krakow, 1989; Cooper, 1991; Browman and Goldstein, 1992; Farnetani and Vayra, 1996); for example, word-initial stops in English have more linguopalatal contact (Byrd, 1994, 1996). For phrase and sentence edges, similar articulatory variation has been found, though the number of studies is smaller: more linguopalatal contact for sentence-initial coronal stops (Keating, 1995), more lip rounding for sentence-initial rounded vowels (van Lieshout *et al.*, 1995). Fougeron and Keating (1996) found less nasal airflow (interpreted as higher velum), and more linguopalatal contact, for French /n/ when initial in a phrase. On the other hand, Byrd *et al.* (1997) found lengthening but no spatial changes at word and phrase edges in Tamil. Finally, acoustic records suggest that glottal articulations also are influenced by phrasal position (Pierrehumbert and Talkin, 1992; Jun, 1993; Dilley *et al.*, 1996).

Another effect on articulation has also been suggested: That articulations are more extreme earlier in utterances and decline¹ gradually over the course of utterances (e.g., Vaissière, 1986; Vayra and Fowler, 1992; Krakow *et al.*, 1994; Hinton, 1996). Vayra and Fowler described the articulatory variation they found for stressed Italian vowels as a “supralaryngeal weakening” and “declination of supralaryngeal gestures” (p. 49), “a weakening of the entire mechanism, respiratory, laryngeal and supralaryngeal for stressing a vowel” (p. 59). Most recently, Krakow *et al.* (1994) found

that the height of the velum during the English /t/ depended on its position from early to late in a sentence. The velum was highest for the earliest /t/, intermediate for a middle /t/, and lowest for the latest /t/. They described “supralaryngeal declination” as “a general ‘winding down’ in speech” (p. 333).

The question arises, then, how supralaryngeal declination, apparently a global effect of serial position in a sentence, is related to the local effect of being at the edge of a prosodic domain. One idea has been that they are the same thing: That domain-final articulations are reduced relative to domain-initial ones precisely because they come later in the domain (Krakow 1989, p. 181). Yet existing data make clear that effects at word and phrase edges cannot be ascribed to simple sentence-level declination, because those effects do not depend on serial order. For example, in Byrd (1994), Jun (1993), and Pierrehumbert and Talkin (1992), the position of test segments in test sentences was controlled across comparisons. If edge effects and declination are to be related, then it must be in a more complicated way.

An important finding about domain-final lengthening is relevant here. Klatt (1975) and Wightman *et al.* (1992) showed that final lengthening is found at more than one domain level, and the lengthening is greater at the end of higher domains than of lower domains. That is, this edge effect operates hierarchically. Similarly, Jun (1993 p. 237) proposed that “there is a hierarchy of strength of prosodic position” to account for her finding that voice onset time (VOT) of a Korean consonant is greatest when phrase-initial, next greatest when word-initial but phrase-medial, and least when word- and phrase-medial. Supralaryngeal declination could also be hierarchically nested [as is f_0 declination (Thorsen, 1985; Maeda, 1976)], and occur not only at the sentence level, but also at the word level, and at phrasal levels in between. Under this interpretation, then, declination would not depend strictly on serial position within a sentence, but instead on serial position within any given prosodic domain.

However, alternative hypotheses to declination are also potentially consistent with previous observations. Figure 1 shows a schematic of different possible patterns within a single prosodic domain, with three points highlighted in each pattern. Articulation varies along some arbitrary dimension on which lower, less extreme, values indicate articulatory reduction. We will refer to less extreme articulations as *weakened* (right panels) and more extreme articulations as *strengthened* (left panels). This is the same notion of weakening as that used by Vayra and Fowler (1992) and in historical linguistics (Straka, 1963). Figure 1(a) and (b) shows progressive trends in the two directions: *Progressive weakening*, corresponding to supralaryngeal declination, and *progressive strengthening*, a kind of reverse declination. In contrast, the other patterns represent more localized effects at domain edges, as has been shown to be the case for domain-final lengthening (Wightman *et al.*, 1992; Beckman *et al.*, 1992). Figure 1(d) shows a pattern that we will call *final weakening*, a local reduction or lenition only at the end of the domain. Figure 1(e) shows the converse, what we will call *initial strengthening*. Initial strengthening accords with

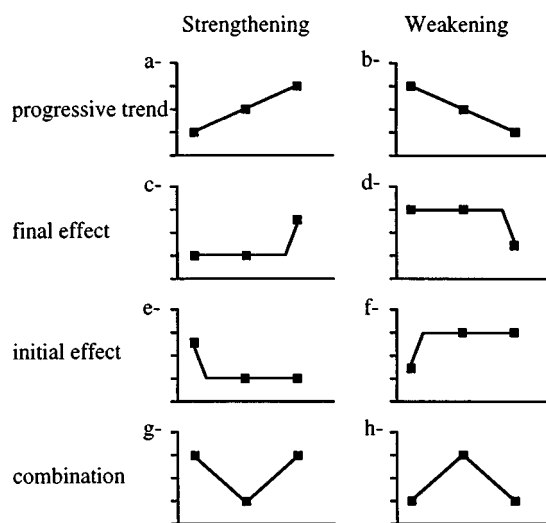


FIG. 1. A schematic of possible patterns of articulatory variation within a prosodic domain. The horizontal axis represents time; the vertical axis represents an arbitrary dimension of articulatory variation, in which lower values are a less extreme articulation. See text for explanation of individual panels.

Fujimura’s suggestion (1990a, p. 232) that “Syllable initial position, as well as word and phrase initial position, seems to be generally characterized by more ‘forceful’ articulatory gestures (...)”. Other possible local effects are given in Fig. 1(f) (*initial weakening*) and (c) (*final strengthening*), and combinations in (g) (V-shaped initial plus final strengthening) and (h) (initial plus final weakening).

Any of these patterns can be hierarchically nested. If some small number of datapoints is then sampled from the whole, then it is possible to obtain a set of three declining datapoints not only from nested progressive weakening, but also from nested initial strengthening or nested final weakening. All that is required is that the test utterance is produced with a certain kind of prosodic organization, in which the hierarchical levels decrease along with serial position in the sentence. That would be any prosodic structure in which the first datapoint is from a segment which is initial in a high prosodic domain (e.g., an utterance), the second segment is initial in some lower prosodic domain (e.g., a phrase), and the third segment is initial in an even lower prosodic domain (e.g., a word). There is thus a distinction between what can be observed in a set of datapoints versus the mechanisms that might underlie and produce that observation. The fact that some measure declines over an utterance does not by itself mean that speech involves a declination mechanism. Therefore our attention is focused on local effects as an alternative to declination.

The present experiment was designed with four goals. The first is to determine whether the articulation of a segment varies depending on its position in long sentences. The second is to determine if any such variation is due to a local strengthening or weakening at particular prosodic positions, or if it is due to a global progressive trend. To do this we compared consonants and vowels in a CV syllable in initial, medial, and final positions within each of four prosodic domains. The third goal is to test whether the articulatory varia-

TABLE I. Four types of models sentences for the reiterant speech with the syllable /no/.

| Model sentences | |
|--------------------|---------|
| 1. 89+89+89+89 | =a lot. |
| 2. (89+89)*(89+89) | =a lot. |
| 3. 89*(89+89+89) | =a lot. |
| 4. (89+89+89)*89 | =a lot. |

tion occurs at more than one prosodic level and is hierarchically cumulative. To do this we positioned test segments at different prosodic levels. The fourth goal is to test for acoustic correlates of these articulatory differences; here we will focus only on durational correlates.

I. EXPERIMENT

A. Method

1. Electropalatography

The articulatory measure for consonants and vowels was linguopalatal contact, contact between the tongue blade/front and the hard palate. Variation in the amount of linguopalatal contact indicates differences in overall oral constriction. Linguopalatal contact was measured by electropalatography (EPG). The Kay Elemetrics Palatometer was used in this experiment; its custom-made pseudopalates have 96 electrodes covering the entire hard palate and the inside surfaces of the molars. Each sweep of the 96 electrodes takes 1.7 ms, and the sampling interval is 10 ms.

2. Test sentences and reiterant speech

To obtain different phrasings in our test stimuli, we used arithmetic statements in which the phrasing of the words would be crucial to conveying the meaning, shown in Table I. The last three test sentences require some prosodic disambiguation, while the first sentence is less constrained. As shown in Table I, these sentences were produced with the numeral “89” (eighty-nine), which usually has lexical stress on its final syllable. Because we were particularly interested in the behavior of the initial syllables, we wanted to avoid confounding stress and positional effects. To control for possible effects of lexical stress, lexical stress location was varied by using two additional trisyllabic numerals, “70” (seventy) with stress on its first syllable, and “100” (one-hundred) with stress on its second syllable.

The test sentences in Table I served as models for reiterant speech, in which each syllable of a model sentence is replaced by a single syllable (here, /no/), but intonation and other prosodic aspects of the model are meant to be preserved. The reiterant utterances used the syllable “no” (/no/), for all of the utterance before “equals.” For example, “(89+89+89)*89=a lot” was read as “(nonono no nonono no nonono) no nonono equals a lot,” giving 15 reiterant syllables. The vowel /o/ was chosen because it is one of the American vowels which occurs in both stressed and stressless syllables, yet is quite different from /n/ in its contact pattern, thus making clear the consonant to vowel difference.

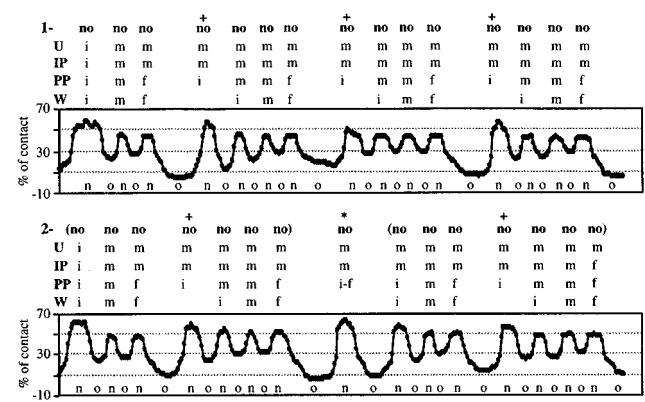


FIG. 2. EPG sample data: The line that moves up and down across the graph is the linguopalatal contact (as percent of electrodes contacted) over time for two samples of the reiterant sentences based on “89+89+89+89=a lot” and “(89+89)*(89+89)=a lot,” produced by speaker 1. At the top of each sample, reiterant syllables (/no/) aligned with percent-contact display, and arithmetic operators + and * indicated; below the percent-contact display, /n/ aligned with maximum contact and /o/ with minimum contact. Above the percent contact display, codings of the 15 test syllables as i(nitial), m(edial), or f(inal) in four prosodic domains: utterance (U), intonational phrase (IP), phonological phrase (PP), and word (W).

3. Subjects and procedure

There were three female American English speakers in the experiment, all phoneticians in the UCLA Phonetics Laboratory. Speaker 1 was the second author. Speakers 2 and 3 were graduate students who had participated in one previous EPG study (Byrd, 1994) but were naive about the present study.

Each test sentence was repeated six times. Speaker 1 read all combinations of the three numerals and four sentence types, for a total of 72 sentences. Speakers 2 and 3 produced all the numerals only in the first sentence type, and the other sentence types only with “89,” for a total of 36 sentences per speaker. Subjects were not told how to phrase each sentence, but instead were simply asked to speak moderately fast and to convey the mathematical meanings indicated, which they all said they understood. Subjects practiced first with the real words, then reiterantly a few times. The audio signal was recorded along with the EPG signal, using an ordinary tabletop microphone in an open laboratory room. Both the audio (12.8-kHz sampling rate, 16-bit resolution) and the EPG signal (100-Hz sampling rate) were recorded digitally in Kay Elemetrics’s Computerized Speech Lab (CSL).

4. Measurements

Linguopalatal contact was measured for each /n/ and /o/. EPG data were analyzed outside CSL by computing the percent of the 96 electrodes contacted in each data frame (Byrd *et al.*, 1995). Each percentage point, then, is about equal to one electrode contacted. Figure 2 shows data for sample tokens by speaker 1, for the first two sentence types, using the number “89” as the model. Notice that in both sample sentences the first /n/ has noticeably greater contact that does the second /n/, while the third /o/ has noticeably less contact than does either of the first two /o/’s.

For each segment (stop or vowel) contact was measured in a single frame, the one showing the extreme contact for that segment. This extreme was defined differently for the /n/ and the /o/, because their articulatory targets are in opposite directions. Because a canonical stop /n/ is characterized by a linguopalatal closure, the target of its articulatory movement was considered to be a maximum contact. Because a canonical vowel /o/ is characterized by a lingual lowering and backing, the target of its articulatory movement was considered to be a minimum contact. With two segments per /no/ syllable, 15 syllables per utterance, and 144 utterances, there was a total of 4320 contact measurements.

In addition, acoustic durations of /n/'s and /o/'s were measured from computer-displayed spectrograms. Segmentation was based on the presence of energy in the region of the higher formants.

Finally, a tonal transcription was made of each token with the help of a trained transcriber who listened to the audio signal and looked at the *f*₀ contour, the segmental durations, and any pauses. Phrase tones and boundary tones were transcribed using the tonal part of the ToBI system (Silverman *et al.*, 1992).

Statistical analyses were performed on the coded data using StatView (Abacus Concepts, 1992). The tests used are described in each results section below. We note here only that, because the speakers selected their own phrasings, and because the categories form a hierarchy, the sample sizes of the different prosodic categories varied widely, but were never smaller than 25 per cell.

5. Prosodic coding

In order to compare different prosodic positions, each segment was coded according to its position in a prosodic structure derived from prosodic hierarchy theory (see Wightman *et al.*, 1992 or Shattuck-Hufnagel and Turk, 1996 for reviews). (1) Our highest level is an utterance (henceforth utterance or U). An utterance here corresponds to a complete sentence. Although not all researchers agree on the existence of the utterance as a distinct prosodic domain, it seems possible that the very beginning of an utterance (in a pretheoretical sense, meaning simply when a speaker begins talking) could have some special status with respect to articulation. (2) This utterance domain contains one or more intonational phrases [henceforth IP, also called Full Intonational Phrases by Beckman and Pierrehumbert (1986)]. An IP is defined by a complete intonational contour, including a final boundary tone. (3) Our intermediate level is a smaller phrasal domain, defined by at least one pitch accent and a phrase tone. It is the Intermediate Phrase of Beckman and Pierrehumbert (1986), and is roughly equivalent to the phonological phrase of other authors. For clarity of abbreviation we will refer to it here as the phonological phrase (henceforth PP). (4) The phonological word (henceforth word or W) in this study corresponds to a numeral like "eighty-nine." While the subparts of "89" and "100" ("eighty," "nine," "one," "hundred") are themselves lexical words, "89" and "100," like other compounds, function as single prosodic as well as

morphosyntactic words. (5) We considered a domain lower than the word, the syllable (henceforth syllable, syll, or S) solely in our across-domain comparisons.

Given the strictly hierarchical relation of these levels, every syllable in every token could be coded according to its position in each domain. Figure 2 shows the coding of the syllables in the two sample tokens. In every case the entire sentence was taken to be a single utterance. This means that the first syllable in each token was coded as U-initial, and all other test syllables in each token were coded as U-medial. There were no U-final test syllables, because the "equals a lot" portion of the tokens was not analyzed. Each numeral was a phonological word, and the initial, medial, and final syllables of the numerals were coded as word-initial, word-medial, and word-final, respectively. For the cross-domain analysis the W-medial syllables were recoded as syll-initial. The arithmetic operators "plus" and "times" were not used as words, since their single syllables are both W-initial and W-final, but they were coded for all higher levels.

Figure 2 also shows codings for the two intermediate prosodic levels used in this study, IP and PP. For these two levels, there was inter- and intra-speaker variation in phrasing (as in Fujimura, 1990a), so that the coding depended on the tonal transcription of each token. Consider first the IP level. Because the prosodic domains are taken to be strictly layered, the first syllable of each token is necessarily initial in an IP. The end of the IP is defined by the presence of a boundary tone. Any syllable after an IP-final syllable is IP-initial; all other syllables are IP-medial. Some tokens, like sample No. 1, had no syllables coded as IP-final, meaning that the first IP continued past the last test syllable into the "equals a lot." In general, there was a tendency for parentheses to delimit IPs, though not in either of the examples shown here. Finally, consider the PP level. The first syllable of a token is necessarily PP-initial. The end of a PP is defined by an intermediate break, and any following syllable is PP-initial. The PPs seen in Fig. 2 comprise either single numerals, or a "plus" and a following numeral, or a "times" alone. Although these are the general tendencies, other phrasings occurred as well. Both operators and numerals were coded for position in these two domains (as well as in the utterance). If an operator formed a PP or IP by itself, it was coded as both initial and final, as with the "PPi-f" in sample No. 2, and was excluded from all further analyses. Although some prosodic domains are consistently longer than others here (an utterance always contains more than 15 syllables and a word only three syllables), PPs and IPs have variable length. Also word-, PP-, and IP- initial syllables can occur in various position in the utterance, so it is not the case that syllables which are initial in a higher level come earlier in the utterance than syllables which are initial in a lower domain. The consonant and vowel of each syllable are coded the same as the syllable as a whole. To avoid potential confusion with absolute positions in a CV, /o/ in a domain-initial CV syllable will be referred to as "initial-syllable /o/" and /n/ in a domain-final CV syllable will be referred to as "final-syllable /n/." Finally, syllables in numerals were coded as stressed or unstressed based on auditory impressions of each speaker's pronunciation.²

B. Results for linguopalatal contact

1. Within-domain comparisons

We tested for variation in linguopalatal contact of /n/ and /o/ within each prosodic domain in order to determine which, if any, of the mechanisms presented in Fig. 1 is at work in our data (especially, a local effect at domain edges or a global trend). Recall that the dependent variable, the measure of articulatory variation, is the extreme value of linguopalatal contact. For /n/, for which the extreme of contact is the maximum value, strengthening would correspond to an increase in contact. In contrast, for /o/, for which the extreme of contact is the minimum value, strengthening would be a decrease in contact. Weakening will show the reverse patterns. A decrease of contact for /o/ can result from a backing or lowering of the tongue, but to simplify here we will describe it as “opening.”

a. Tests for domain initial effects: initial versus medial and final syllables. In this analysis we compare the initial syllable with the medial and final syllables in each prosodic domain to test for local effects in domain-initial position. First, U-initial segments were compared to U-medial segments, i.e., all other segments (recall that there are no U-final test segments). Second, IP-initial segments which were not also U-initial were compared to IP-medial and IP-final segments. Third, PP-initial segments which were not also IP-initial or U-initial were compared to PP-medial and PP-final segments. Finally, W-initial segments which were not also PP-initial, IP-initial, or U-initial were compared to W-medial and W-final segments. Therefore, in these analyses, “IP-initial,” for example, means “the highest domain in which this syllable is initial is the IP.” This limitation to “exclusive” coding was necessary to ensure that initial segments in some smaller domain would not have greater average contact than the medials or finals because some initials were also initial in a larger domain. With three speakers, there were thus a total of 12 comparisons.

Figure 3 shows the average values for each speaker of these linguopalatal contact measures for the /n/’s and the /o/’s in initial, medial, and final positions within each prosodic domain. For this analysis, all data from the three subjects were used. Table II gives the results of statistical comparisons by one-factor analysis of variance (ANOVA) and Fisher’s protected least significant difference (PLSD) *post hoc* pairwise comparisons, with 0.05 as the significance level for all tests. The null hypothesis is that there is no difference in contact across these positions within each domain.

Consider first the /n/’s. For all three speakers, U-initial /n/’s have significantly more contact than U-medial /n/’s; IP-initial /n/’s have significantly more contact than IP-medial and IP-final-syllable /n/’s; PP-initial /n/’s have significantly more contact than PP-medial and PP-final-syllable /n/’s. More linguopalatal contact in initial position for /n/ is an initial strengthening. On the other hand, no general pattern is seen within words; only speaker 3 has more contact for W-initial /n/ than for both W-medial and W-final-syllable /n/. Speaker 1 does not show variation across word positions. For speaker 2, Word initial /n/’s have more contact than medial ones but not more than final ones.

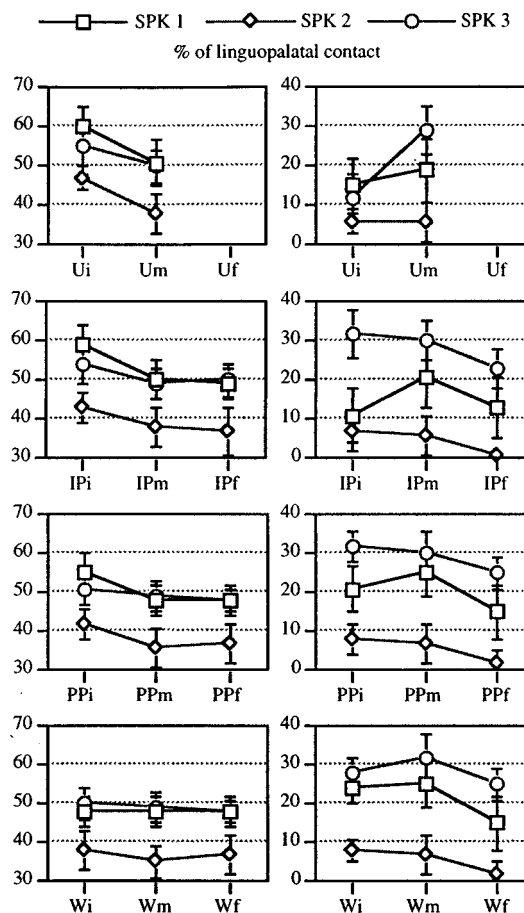


FIG. 3. Maximum linguopalatal contact for /n/’s (left) and minimum linguopalatal contact for /o/’s (right) in three positions (initial, medial, final) in each of the four prosodic domains (utterance, intonational phrase, phonological phrase, word). Speaker results are shown separately within each panel. See Table II for significance of comparisons. All data from all speakers are included here, coded exclusively. A more extreme articulation is more contact for /n/ and less contact for /o/.

Note that this initial strengthening holds although the prosodic coding is exclusive for domain initial /n/’s. However, another possible confound in this kind of analysis could be that the higher-domain-medial /n/’s would have less contact simply because they would also be in medial or final syllables in the word. Nonetheless, when we compared initials, medials, and finals within the PP and IP levels, taking only /n/’s which are W-initial or in operators, much the same result was found [(b) in Table II]. IP-initial /n/’s still had significantly more contact than medials for all three speakers. PP-initial /n/’s still had significantly more contact than medials for two of the three speakers (speakers 1 and 2); for speaker 3 (the subject who has more contact W-initially than medially and finally) the direction of difference was maintained but it was no longer significant. Thus when the word and PP levels are taken together, every speaker distinguishes initial consonants in one or the other domain.

Another confounding factor could be the position of lexical stress—this could favor one possible mechanism compared to the others, especially at the word level. For speaker 1, we recorded the reiterant version of the three numerals in all the sentence types. When the three lexical stress

TABLE II. Results of statistical comparisons for Sec. I B 1. Fisher's PLSD *post hoc* comparison of percent of linguopalatal contact between initial versus medial versus final /n/'s and /o/'s at the four prosodic domains defined for the three speakers. * = $p < 0.05$; ns = $p \geq 0.05$. The columns for each domain correspond to: (a) all data in Fig. 5, (b) only W-initial and operators, (c) not W-initial or PP-initial.

| All numerals | /n/ | | | | | | | | /o/ | | | | | | | |
|-----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | U | | IP | | PP | | W | | U | | IP | | PP | | W | |
| Speaker 1 | (a) | (a) | (b) | (c) | (a) | (b) | (c) | (a) | (a) | (a) | (b) | (c) | (a) | (b) | (c) | (a) |
| initial versus medial | >* | >* | >* | | >* | >* | | ns | <* | <* | <* | | <* | <* | | <* |
| initial versus final | | >* | | | >* | | | ns | | ns | | | >* | | | >* |
| medial versus final | | >* | | ns | ns | | ns | ns | | >* | | >* | >* | | >* | >* |
| Speaker 2 | | | | | | | | | | | | | | | | |
| initial versus medial | >* | >* | >* | | >* | >* | | >* | ns | ns | ns | | ns | ns | | >* |
| initial versus final | | >* | | | >* | | | ns | | >* | | | >* | | | >* |
| medial versus final | | ns | | ns | ns | | <* | <* | | >* | | >* | >* | | >* | >* |
| Speaker 3 | | | | | | | | | | | | | | | | |
| initial versus medial | >* | >* | >* | | >* | ns | | >* | <* | ns | ns | | >* | >* | | <* |
| initial versus final | | >* | | | >* | | | >* | | >* | | | >* | | | >* |
| medial versus final | | ns | | <* | ns | | ns | ns | | >* | | >* | >* | | >* | >* |

patterns are equally represented in this way, initial strengthening of /n/ is found at every level except the word level, suggesting that lexical stress is not the cause of the overall initial strengthening pattern. However, at the word level no distinction is found between initial, medial, and final-syllable /n/'s, suggesting that there could be an effect of lexical stress.

Next consider the /o/'s, shown in the panels at the right in Fig. 3. Recall that initial strengthening, a more extreme articulation in domain initial position, would mean for /o/ less contact (greater opening), while initial weakening would mean more contact. Initial-syllable /o/'s (/o/'s in the initial CV syllable of a domain) have less contact than medial /o/'s for only half of the 12 comparisons: within all levels for speaker 1, and within utterance and word for speaker 3. In two comparisons initial-syllable /o/'s have more contact than medials, and in the other comparisons they have the same contact. So for the vowel /o/, there is some initial strengthening, but it is less prevalent than that observed for /n/, being consistent for only one speaker. These results hold whether we include only vowels in the word-initial syllable, or all

vowels, at higher prosodic levels [Table III(a) and (b)].

b. Tests for domain-final effects: final versus initial and medial syllables. In this analysis we compare the final syllable with the medial and initial syllables in each prosodic domain to test for local effects in domain-final position. This test cannot be done at the utterance level, since there are no U-final test syllables. With three speakers and three levels, then, there are nine tests to be made for /n/ and for /o/. First, consider the consonants. (Recall that final-syllable /n/'s are the /n/'s in the final CV syllable of a domain.) In Fig. 3, the difference between medial and final-syllable /n/'s can be seen to vary in size and direction across levels and speakers; sometimes medials have more contact than finals (for example, speaker 1 in the IP). However, again we must be careful about confounds across levels. Medials could have more contact than finals simply because more of the medials could be initial in some lower domain(s) and get strengthened at that lower level. This is likely to be the case because domain-final syllables are almost never initial in lower domains. Therefore only domain-medial and domain-final-syllable /n/'s which are not initial in any lower domain (word

TABLE III. Hierarchical prosodic levels significantly ($p < 0.05$) distinguished by the amount of (a) linguopalatal contact for initial /n/; (b) linguopalatal contact for the final vowel V1 preceding initial /n/; (c) linguopalatal contact for the initial-syllable vowel V2 following initial /n/; (d) C-to-V contact difference (/n/ minus V2); (e) V-to-C contact difference (/n/ minus V1); (f) acoustic duration of /n/; (g) acoustic duration of final /o/ (V1). Results presented by speakers for the subset of data with the numeral "89."

| | Speaker 1 | Speaker 2 | Speaker 3 |
|-----|--------------------------------------|------------------|----------------------------------|
| (a) | IPi>PPi>Wi=Si (Ui=PPi and Ui=IPi) | Ui>IPi=PPi>Wi>Si | Ui=IPi>PPi=Wi>Si |
| (b) | IPf<PPf<Wf=Sf | IPf=PPf<Wf=Sf | IPf,PPf,Sf<Wf |
| (c) | IPi<Ui<PPi<Wi<Si | Ui,IPi,Si<PPi | Ui<Wi<PPi=IPi<Si |
| (d) | IPi>Ui>PPi>Wi>Si | Ui>IPi>PPi>Wi>Si | Ui>IPi=PPi>Si (Wi=IPi,Wi>PPi) |
| (e) | IPi>PPi>Wi=Si | IPi=PPi>Wi>Si | IPi=PPi=Si>Wi (IPi>Wi) |
| (f) | IPi>PPi>Wi>Si=Ui | IPi=PPi>Wi>Ui>Si | Ui=IPi>PPi>Wi>Si |
| (g) | IPf=PPf>Wf=Sf | IPf=PPf>Wf=Sf | IPf>PPf>Sf>Wf |

or PP) were compared [(c) in Table II, and column (a) for the word level]. Of the nine comparisons, three give a significant difference between medial and final-syllable /n/'s: at the word and PP levels for speaker 2, and at the IP level for speaker 3, with final-syllable /n/'s having more contact than medial /n/'s, i.e., final strengthening. For the other six cases final-syllable and medial /n/'s have the same contact.

For the vowels, it can be seen in Fig. 3 that /o/'s in domain-final positions generally show the least contact. Domain-final /o/'s are more open than medial /o/'s at every level for all speakers and domain-final /o/'s are usually more open than /o/'s in domain-initial syllables, with speaker 1's IP domain the only exception. Greater opening or backing is a more extreme articulation for /o/, thus a final strengthening. This result cannot be due only to lexical stress on the final syllable of "89," because in the data for speaker 1 three lexical stress patterns are included equally, and final strengthening is the same for this speaker as for the other two.

c. Tests for declination within-domain and across serial position. In the previous sections we have shown that initial /n/'s are more extreme in their constriction than medial and final-syllable ones, and final /o/'s are more extreme in their opening than initial and medial ones. There is some initial strengthening of the /o/'s in the initial CV syllable, and less consistent final strengthening of the /n/'s in the final CV syllable. In this section, we compare these results to see if they form a global trend across the three positions. In Fig. 3, a declining pattern will show a progressive decrease of contact for /n/'s (as they become less closed), and a progressive increase of contact for /o/'s (as they become more closed).

For /n/, the usual pattern is simple initial strengthening [Fig. 1(e)], but there are three comparisons showing initial strengthening combined with final strengthening [Fig. 1(g)]. There are no cases of declination. Thus the initial strengthening of /n/ is not part of a larger declining trend. For /o/, the relation across the three positions shows three different patterns. Four comparisons show a combination of initial and final strengthening [Fig. 1(g)], and three show simple final strengthening [Fig. 1(c)]. Two comparisons show a progressive opening trend [Fig. 1(a)], with final /o/'s more open than medial /o/'s and medial /o/'s more open than initial-syllable /o/'s. This progressive opening of /o/ is significant for speaker 2 at the word level, and for speaker 3 at the PP level. Thus, for these /o/'s we observed not a declination but instead a progressive strengthening. In sum, evidence for a within-domain declination has been found for none of the consonant comparisons; two vowel comparisons show a reverse declination, i.e., progressive strengthening.

We also tested for a sentence-level global trend that would depend only on serial position of a segment in the sentence. We tested whether the amount of linguopalatal contact varies linearly, either over the whole sentence or within smaller domains. This was done by testing for correlations between linguopalatal contact and linear position in the sentence, regressing serial position of the consonants (from syllable 1 to syllable 15) against their amount of contact, and using all data from all speakers. In the previous analysis we averaged all domain-medial syllables and com-

pared them as a group to averaged domain-initial ones and domain-final ones; here we code all syllables by serial position so each could show its contribution to a potential progressive trend. We found no such trends for /n/'s or /o/'s ($r^2 \leq 0.01$) for all comparisons: over /n/'s in all syllables; stressed syllables; W-initial syllables; PP-initial syllables; IP-initial syllables; over /n/'s only in medial and final syllables, taking out the strengthened W-initial, PP-initial, IP-initial, and U-initial syllables; over all /o/'s or (following Vayra and Fowler, 1992) stressed /o/'s only). Thus we can say that in our data there is neither overall declination nor overall progressive strengthening for either consonants or vowels.

2. Hierarchical level of the domain boundary (across-domain comparisons)

a. Effects on linguopalatal contact. The next analysis focuses on the hierarchical nature of the strengthenings found above. For example, we ask whether the initial strengthening found for /n/ in different prosodic domains is cumulative, whether contact for initial consonants is greater when they are initial in higher prosodic domains than when they are initial only in lower prosodic domains. The comparisons were made by ANOVA followed by Fischer LPSD *post hoc* comparisons, with 0.05 as the significance level, for all pairings of domain initial /n/'s: W-initial /n/'s which are not also initial in PP, IP, or U; PP-initial /n/'s which are not also initial in IP or U; IP-initial /n/'s which are not also initial in U; and U-initial /n/'s. In order to compare W-initial consonants with ones initial in a lower domain, we included in the comparison syll-initial consonants that are not also W-initial or W-final (i.e., W-medial consonants). Similar comparisons are made with domain-final vowels and vowels in domain-initial syllables.

Figure 4(a) shows the average maximum contact for initial /n/'s according to the consonants' highest domain, for just the sentences with "89" (the numeral for which all the speakers produced all the sentence types) for the individual speakers. The general tendency is that the contact is related to the hierarchical level of the domain boundary: Higher levels show more contact. However, as can be seen in the figure, speakers vary. Not all levels are reliably distinguished, and although three or four levels are distinguished, the speakers differ in which levels those are. Table III(a) summarizes the significant differences for each speaker. In speaker 1's data, IP-initial, PP-initial, and W-initial/syll-initial /n/'s are distinct, but U-initial is not different from either IP-initial or PP-initial, and W-initial is not different from syll-initial (W-medial) (as already seen in the previous section). Speaker 2 distinguishes U-initial, IP-initial/PP-initial, W-initial, and syll-initial, but IP-initial is not different from PP-initial. Speaker 3 distinguishes IP-initial/U-initial, W-initial/PP-initial, and syll-initial, but U-initial and IP-initial are not different from each other, nor are PP-initial and W-initial (as already seen in the previous section). In sum, syllable is distinguished from word (2 speakers), word is distinguished from PP (2 speakers), and PP is distinguished from a higher domain (3 speakers), but whether that higher domain is the IP or the utterance is variable.

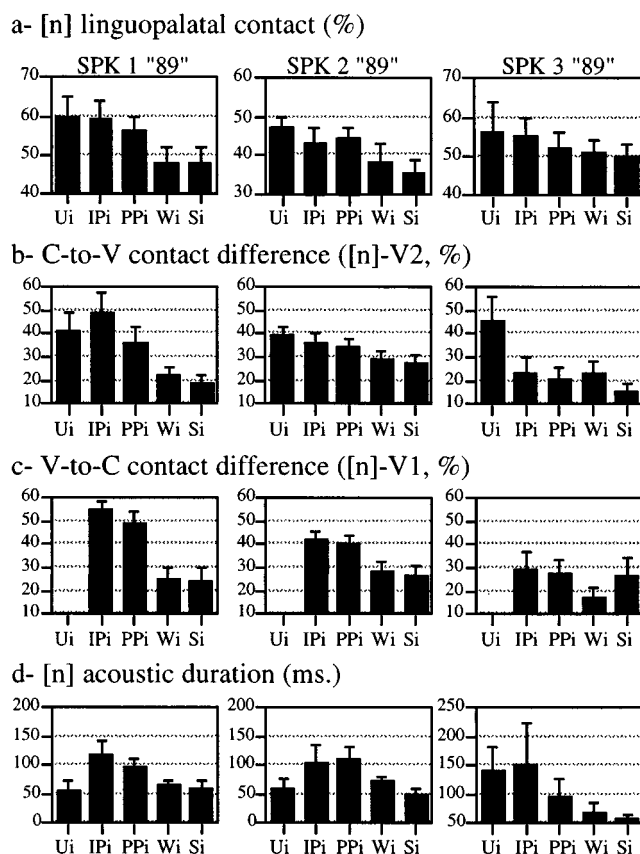


FIG. 4. Data on initial position at five hierarchical levels (utterance, intonational phrase, phonological phrase, word, syllable) for (a) maximum linguopalatal contact for /n/; (b) difference in contact between initial /n/ and following vowel (C-to-V); (c) difference in contact between initial /n/ and preceding (final) vowel (V-to-C); (d) acoustic duration for /n/. Data for three speakers, subset of the corpus with "89" in four sentence types.

We can also ask whether the strengthening found for domain-final /o/'s is cumulative in the same way. As there are no U-final /o/'s, there are four levels that could be distinguished. Some tokens had to be excluded from this analysis because an /o/ showed no contact at all: Its opening is greater than that of /o/'s showing contact on the pseudopalate, but we have no way to infer its real degree of opening. There were 47 such /o/'s excluded, generally IP-final, sometimes PP-final, mostly for speaker 2. Table III(b) presents the comparison for the domain-final vowels (V1) preceding the strengthened initial /n/. No speaker distinguishes all four levels. The only common result is that W-final /o/'s are consistently less open than PP- and IP-final /o/'s. It seems that above the word, final /o/'s are simply always quite open, their degree of openness depending very little on their hierarchical prosodic position.

Recall that speaker 1 and to some extent speaker 3 showed strengthening of /o/'s in domain-initial syllables (V2). We can ask whether this strengthening is cumulative too. Results are presented in Table III(c). More levels are distinguished by /o/'s in initial syllables than by final /o/'s, but the number and order are different for the speakers. The only common result, which in fact holds for all three speakers, is that U-initial-syllable /o/ has less contact (more strengthening) than PP-initial-syllable /o/.

b. Effects on V-to-C and C-to-V linguopalatal contact difference. Above we have seen the effects of prosodic position on the extreme contact for /n/ and for /o/, considered separately. In this section we consider the entire sequence V_1CV_2 , to see whether the articulatory contrast between successive segments is also affected by their prosodic position. One measure of this is the contact difference between the preceding /o/ (V1) and /n/, the V-to-C contact difference. These segments are heterosyllabic and span a prosodic boundary. The other measure is between the /n/ and the following /o/ (V2), the C-to-V contact difference. These segments are tautosyllabic and the syllable follows a prosodic boundary.

Figure 4(b) shows the C-to-V contact difference for the subset of the corpus with "89" for the three speakers. Significance of the comparisons is presented in Table III(d). The C-to-V contact difference generally shows more distinctions than either /n/ or /o/ alone—five levels for two speakers, three for the third—and in general these distinctions follow the same hierarchical order as do those made by /n/ contact. Figure 4(c) and Table III(e) show the V-to-C contact difference and the significance of the comparisons. This difference is large, but not strongly cumulative.

Another way to see if an increase in contact for /n/ is accompanied by a change in contact for /o/ is to look for a correlation. For speaker 1, a strengthened /n/ is also surrounded by strengthened (i.e., more open) vowels, as shown by a negative correlation between the amount of contact for /n/ and the amount of contact for V1 ($r^2=0.5$) and for V2 ($r^2=0.2$). For speaker 2, a strengthened initial /n/ is accompanied by an increase of contact of the following vowel (V2 less open, positive correlation, $r^2=0.2$), while the amount of contact of the preceding vowel is not correlated with the contact of /n/. For speaker 3, the contact of vowels and /n/'s is independent ($r^2=0.01$).

3. Results for acoustic duration

Figure 4(d) presents the average acoustic duration of initial /n/'s according to the consonant's highest domain, for the subset of the corpus with "89" for the three speakers. Significant comparisons are given in Table III(f). Comparison with Fig. 4(a) shows that the durations of /n/ follow the same patterns as for linguopalatal contact, and except for U-initial /n/, they follow the prosodic hierarchy, with greater initial lengthening in higher domains. Speakers 1 and 3 distinguish four prosodic levels by /n/ duration (IP-initial greater than PP-initial, greater than W-initial, greater than syll-initial). The acoustic duration of U-initial /n/ is like syll-initial /n/ (speaker 1), or IP-initial /n/ (speaker 3). Speaker 2 does not distinguish between IP-initial and PP-initial /n/'s, and U-initial /n/'s are shorter than W-initial. Thus speakers 1 and 3 each make one more prosodic-domain distinction by /n/ duration than by /n/ linguopalatal contact, and speaker 2 the same number.

In contrast to the results for initial /n/, domain-final vowel lengthening [Table III(g)] is only weakly cumulative. For two speakers (1 and 2) only two levels are distinguished by duration, IP-final and PP-final being longer than W-final and syll-final /o/'s. For speaker 3, four levels are distin-

guished, but not in the expected order: IP-final is longer than PP-final, which is longer than syll-final, which is longer than W-final. Thus, in our corpus, final vowels are generally poor indices of the hierarchical level of prosodic domains both in their spatial (linguopalatal contact) and temporal characteristics, though all three speakers have longer vowels phrase-finally than W-finally.

Since duration and linguopalatal contact pattern similarly, linear regressions of acoustic duration of /n/ against maximum linguopalatal contact were carried out. For all speakers, all segments in the four sentence types with “89,” and in the flat sentence type with “70” and “100,” were used (this is more of the data than used in the factorial analyses reported above). Despite the similarity of patterning for the two variables, the correlation between them was minimal, whether we consider all /n/’s ($r^2=0.06$), domain-initial ($r^2=0.04$) or final-syllable /n/’s ($r^2=0.006$), or all the /n/’s except initial ones ($r^2=0.05$) or except final-syllable ones ($r^2=0.09$) or except the U-initial ones ($r^2=0.07$) (which show the greatest variation between speakers). The vowels, on the other hand, show a somewhat stronger correlation—the longer the vowel is, the less contact it has (the more open it is) ($r^2=0.3$). So for vowels, duration may be a contributing factor to the degree of opening (or the reverse). This correlation would likely be greater if the most open vowel positions were more reliably tracked by EPG.

II. GENERAL DISCUSSION AND CONCLUSION

We found that the articulation of consonants and vowels varies as a function of their position in long sentences. This variation appears to be a localized effect at prosodic domain edges, i.e., a strengthening of initial consonants and final vowels, and not a global declining trend. Initial strengthening of consonants is found at different prosodic levels and tends to be cumulative. Finally, this cumulative initial strengthening of consonants is accompanied by a cumulative lengthening, though the correlation between strengthening and lengthening is weak.

A. Initial strengthening

1. Within domains

Within each prosodic domain, /n/’s in initial CV syllables were found to have greater contact than /n/’s in medial and final CV syllables, while /n/’s in medial and final CV syllables had comparable contact, at the utterance, IP, and PP or word levels. That is, we found initial strengthening for /n/ in that the articulation of /n/ was more extreme in domain-initial positions. There were three cases combining initial and final strengthening of /n/. There was no declination in the articulation of /n/ at any level for any speaker. Neither was there final weakening; however, it must be borne in mind that the possibility of a weakening of final (coda) consonants at any or all prosodic levels cannot be rejected, since we looked only at CV syllables.

When the same comparisons were made for /o/’s, the only consistent result was that /o/’s in final syllables were found to have less contact than /o/’s in initial and medial syllables in almost every case. Since less contact means

greater opening/backing for /o/, this is a more extreme articulation and therefore a final strengthening. We also observed some combinations of final and initial strengthening, and some progressive strengthening, but because these were not consistent, no reliable overall pattern across the three positions emerges for /o/ in the way that one does for /n/.

Finally, we showed that most of the prosodic domains considered are delimited by strengthened articulations. The beginning edge of prosodic domains is marked by lengthening and increasing contact for the consonant. The final edge of prosodic domains is marked by lengthening and decreasing contact for the vowel. As a result of these initial and final edge effects, a prosodic domain is set off at both its edges. This finding can be related to Dilley *et al.* (1996), who showed that glottalization of word-initial vowels occurs more frequently at onsets of higher prosodic constituents, therefore marking off these constituents.

2. Across domains

When /n/’s that are initial in different prosodic domains are compared, initial strengthening is found to be somewhat cumulative. /n/’s that are initial in some higher prosodic domains have more linguopalatal contact than /n/’s initial in some lower prosodic domains. The experiment allowed comparisons of five levels of prosodic structure: syllable, word, phonological (or intermediate) phrase, intonational phrase, and utterance. Overall, three or four levels were distinguished by significant differences in linguopalatal contact of the initial consonant. IP-initial /n/’s were always distinct from word- and syllable-initial /n/’s. But differences between speakers are noteworthy, in both how many, and which, levels were distinguished. Thus while there is clearly some cumulative initial strengthening across domains, it does not seem to be tied to distinguishing specific prosodic levels; we can say only that speakers distinguished three or four levels this way.

There is no reason to expect any one articulatory or acoustic correlate to distinguish all prosodic domains. Wightman *et al.* (1992) found that preboundary vowel duration distinguished only four of the seven perceptually distinct levels they tested. If we compare the distinctions made by acoustic duration and by different linguopalatal contact measures, it seems that C-to-V difference and initial lengthening are, for two of the three speakers, better correlates of the prosodic hierarchy. In contrast, it is surprising that final vowel lengthening marked only two distinctions in our corpus, between word and higher levels, but not between the two phrasal levels as was shown in Wightman *et al.* (1992). We have no explanation for this.

3. What about declination?

In the Introduction we sketched an alternative account of supralaryngeal declination of initial consonants as described by Krakow *et al.* (1994), based on nested initial strengthening or final weakening. Our experiment shows initial strengthening for consonants; can we then account for declination by this strengthening? The plausibility of an initial strengthening account of declination over three datapoints

depends on whether initial strengthening can distinguish three prosodic levels. Initial strengthening would give a pattern looking like declination if the consonants compared are initial in these three prosodic domains and if the hierarchical levels of these domains decreases along with their serial position in the sentence. To see this, consider again Fig. 2, our sample data, for example the top utterance. It is possible to pick three datapoints for /n/ which show declination: for example, the first /no/ (which is U-initial and IP-initial), the eighth /no/, corresponding to the second “plus” (which is PP-initial), and the 13th /no/, corresponding to the beginning of the last numeral (which is W-initial). This apparent declination is only the consequence of three different degrees of strengthening at the word, PP, and IP levels. If only these points had been considered, we would have concluded that there is declination. Yet all of our systematic comparisons testing for declination yielded none.

B. What is the nature of this strengthening?

The terms “strengthening” and “weakening,” as well as the variants “fortition” and “lenition,” are often used to characterize segment variation or historical changes (Hock, 1992), but it is seldom that those terms are phonetically or articulatorily defined, or that the mechanism leading to the variation is explained. Our results lead us to think that the variation we observe is the result of a general phenomenon in speech, which we call *articulatory strengthening at prosodic-domain edges*. We have considered this strengthening to mean more extreme articulation, that is, spatial variation. Strengthening may also involve greater lengthening, that is, temporal variation. Our observations of linguopalatal contact variation are only the result of this strengthening and are not sufficient to establish its nature. Here we will propose and discuss some possible mechanisms that may induce the more extreme articulations of strengthened segments.

(1) *Increased duration*. In general, shorter durations often (though not necessarily) lead to articulatory undershoot (Lindblom, 1963; Moon and Lindblom, 1994). Conversely, stronger segments could have more extreme articulations because they are longer, and thus have time to reach their targets. However, in our data the spatial variation observed is not strongly correlated with the temporal variation.

(2) *Increased distance between segments*. Recall that there is greater difference in linguopalatal contact between /n/ and /o/ at higher prosodic boundaries, for V-to-C and especially for C-to-V. Possibly, after an open final vowel there could be overshoot of the following lingual target. This was suggested for the jaw in /ata/, /asa/, and /ada/ (though not /ana/) sequences by Keating *et al.* (1994), or, since movement velocity is usually proportional to displacement, the larger displacements from an open final vowel to the following initial consonant would involve higher velocities. Higher velocity would result in a greater impact of the tongue against the palate at closure (A. Löfqvist, personal communication). In both cases (overshoot or higher velocity) more compression of the tongue tissue, and therefore greater contact, would result. But in both cases if it is /n/’s distance from the preceding vowel that leads to /n/’s greater contact, then we should see a negative correlation between the con-

tact for V1 and C. This is so for speaker 1, C has the most contact just when V1 has the least. And speaker 1 is the one subject who has consistent initial strengthening of /o/, V2 has the least contact just when C has the most. So for this subject, the spatial distance between successive oral targets could be an important factor in initial strengthenings. Whether the relevant distances, movements, and velocities are those of the tongue, the jaw, or both, cannot be addressed with EPG.

(3) *Increased coarticulation*. Strengthening could increase the overlap of /n/ with surrounding vowels, in which case contact at the back of the palate for /o/ could occur during /n/, increasing the total contact measured for /n/. In this case there should be some correlation between the contacts for adjacent /n/’s and /o/’s, such that when one has more contact the other also has more contact. For one speaker (speaker 2) there is a weak correlation of this kind ($r^2=0.2$); for the other speakers the correlations were either near zero, or showed less coarticulation. In general these correlations and the increased displacement between C and V’s at higher prosodic boundaries do not support this hypothesis. Furthermore, this hypothesis would have nothing to say about the decreased contact for strengthened /o/.

(4) *Greater coarticulatory resistance*. Conversely, stronger segments could resist coarticulatory undershoot because they resist blending with overlapping gestures (Fowler and Saltzman 1993, p. 182). Strengthening of /n/ would involve more contact because the tongue blade is not pulled away from its constriction target by /o/’s tongue body articulation, and vice versa for /o/. This hypothesis is supported by the increased V-to-C and (especially) C-to-V displacements at higher prosodic levels.

(5) *Increased effort or energy*. Articulatory strengthening could also result from a greater overall effort in speech that would also affect the pulmonary and laryngeal systems, as has been proposed for stress (Ladefoged, 1967; Sluijter *et al.*, 1997). Variations observed in supralaryngeal articulations would only be an indirect effect of this overall energy increase (Öhman, 1967; Vayra and Fowler, 1992). Another possibility could be that initial strengthening is the result of a localized increase in supralaryngeal articulatory effort. Fujimura (1990a) suggested that prosodic-domain initial consonants are characterized by more “forceful” articulatory gestures. Variation of articulatory effort was also suggested by Straka (1963 p. 91), who defined articulatory energy in terms of the force of contraction of the muscles primarily involved in the articulation of the segment, specifically excluding respiratory and laryngeal systems. Straka found that in more “forceful” (“renforcée”) pronunciation there was greater linguopalatal contact for consonants and less contact for all vowels, and as a consequence an increased difference in openness between successive segments. This is what we found at prosodic domain edges.

In conclusion, there are a number of possible mechanisms that might result in the strengthening we have observed, but resolving this question would require much further work.

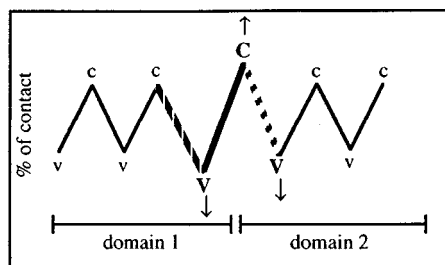


FIG. 5. Schematic summarizing linguopalatal contact for segments spanning a boundary. Dashed line shows difference between final-syllable /n/ and final /o/ in a domain; bold line shows difference between final /o/ and initial /n/; dotted line shows difference between initial /n/ and initial-syllable /o/.

C. Enhancement and the listener

The mechanisms discussed above present strengthening as either the automatic, unplanned consequence of some other aspect of speech production, or something learned as part of the language. In either case it might be useful linguistically. We can think of three ways in which initial strengthening could benefit a listener. Two of these have to do with prosodic parsing. First, strengthening could help with segmentation of the signal into words and higher domains. Recall that in Sec. 1B 2*b* it was shown that there is some enhancement of V-to-C and C-to-V linguopalatal contact differences at prosodic boundaries. Figure 5 combines these two aspects into a single scheme for a CV#CV sequence, where # is some prosodic boundary. This schematic shows that the articulatory contrast between the consonant after the boundary and its surrounding vowels is enhanced because they are more extreme in opposite directions. This articulatory enhancement of the contrasts within the sequence may contribute to marking the prosodic boundary even more clearly than do the vowel or consonant alone. A similar enhancement of CV contrasts at domain-initial boundaries in Italian has been discussed by Farnetani and Vayra (1996, p. 12): "Also boundaries are signalled by an increase in CV contrast: initial boundaries are marked by a strengthening of consonant closure and by an increase in vowel posteriority (...)."

Second, the degree of strengthening could possibly tell the listener about the strength of the prosodic boundary, similar to the way that Wightman *et al.* (1992) suggest that listeners could use degree of final lengthening. A listener would know that when an initial consonant is more than minimally strengthened, the boundary (or break) must be stronger. In particular, IP boundaries could be distinguished from word boundaries in this way. Our results would not support a stronger claim, that listeners might judge the absolute level of any new domain from the degree of strengthening. There is simply too much interspeaker variation.

The third way in which initial strengthening could benefit a listener concerns lexical access. If initial strengthening enhances the segment-specific articulations of consonants and vowels, then it could enhance cues that aid in identifying each segment. de Jong (1995) proposes that stress involves a local hyperarticulation that makes each segment more different from all other segments of the language, so that lexical

contrasts are more distinctive. Along the same lines, enhanced accessibility of segmental information in domain-initial positions would be particularly helpful, particularly for word-initial segments which are important in word recognition (Cole *et al.*, 1978; Hawkins and Cutler, 1988) and at domain beginnings where there is less top-down (e.g., syntactic and semantic) information available.

Of course the linguistic function of initial strengthening presupposes that strengthening has one or more acoustic/auditory correlates. We found that the variation in linguopalatal contact for /n/ and /o/ is accompanied by variation in acoustic duration, though the two measures are not strongly correlated. The acoustic duration differences would be potentially available to listeners. It remains to be seen whether the linguopalatal contact differences have any other associated acoustic properties, and whether these can be heard by listeners.

D. Prosody and articulation

Our results underline a point made by a few other researchers: the importance of understanding, controlling, and reporting the prosody of speech materials in articulation experiments. For the individual experimenter, unsought variation in prosody is a potential confound both within and across speakers, as our own experiment shows. It can also make comparisons across studies difficult or impossible, as researchers have always known. Yet, at the same time, awareness of prosodic differences between sentences can turn apparent random variation into predictable, lawful regularities of speech production, as we hope to have shown here.

This is not to say that prosody is easy to control, or that experiments on prosodic effects are easy to design. It can be especially difficult to find sequences of real words that can occur across a variety of prosodic boundaries and that contain segments appropriate to a given method of articulatory data collection (in our case, lingual consonants). Reiterant speech finesses this difficulty, but some subjects may not be able to produce it fluently (Larkey, 1983) (thought this was not a problem in the present study). It may also be the case that reiterant speech induces somewhat exaggerated rhythmic alternations. These in turn may enhance the prosodic phrasing, and thereby its manifestation in articulation. The present experiment does not address this point, and our results need to be confirmed with experiments using real words (see Fougerson and Keating, 1996 for French).

Another difficulty in such experiments is that the phrasings of the test sentences are unclear unless the utterances are prosodically analyzed. The use of orthographic devices such as punctuation (e.g., commas) or parentheses in the test materials does not guarantee that subjects will produce any particular phrasing. In our study there was variation both within and across speakers, requiring *post hoc* prosodic transcription to determine the actual phrasing of each token.

E. Conclusion

Most previous research on prosodic demarcation has focused on the ends of prosodic domains. Our results add to the much smaller literature on beginnings of domains. At

least some tongue blade and body articulations are more extreme in domain-initial positions. We have shown that position with respect to prosodic boundaries affects articulatory constriction, something that would seem to be an inherent property of a sound.

ACKNOWLEDGMENTS

This work was supported by an MRT allocation to the DEA de Phonétique de Paris to the first author and by NSF Grant No. SBR-9511118 to the second author. Earlier versions were those presented at the Spring 1995 meeting of the Society in Washington DC, and appearing in UCLA Working Papers in Phonetics 92 (1996). We wish to thank the subjects (Kimberly Thomas and Barbara Blankenship) for participating; Sun-Ah Jun and Sabrina Cooper for help with data analysis; Janet Pierrehumbert for providing information about the sentences in Pierrehumbert and Talkin (1992); and Dani Byrd, Edda Farnetani, Bruce Hayes, Sun-Ah Jun, Rena Krakow, Stefanie Shattuck-Hufnagel, Jacqueline Vaissière, John Westbury, Jie Zhang, Editor Anders Löfqvist, Janet Fletcher, and two anonymous reviewers for helpful comments.

¹The term "declination" originally comes from the study of intonation, where it refers to a downtrend in fundamental frequency (f_0) [see Ladd (1984) or 't Hart *et al.* (1990) for reviews]. Subglottal pressure also declines over an utterance, and some of the downtrend in f_0 has been attributed to the downtrend in subglottal pressure (Gelfer *et al.*, 1987).

²The first syllable of "70" and the second syllable of "100" were stressed for all three speakers, and other syllables in those numerals were unstressed. For two speakers, the last syllable of "89" was stressed and the other syllables unstressed, but for one speaker (speaker 3) both the first and last syllables of "89" were stressed and only the middle syllable was unstressed.

Abacus Concepts (1992). *StatView* (Abacus Concepts, Inc., Berkeley, CA).
 Beckman, M., and Edwards, J. (1994). "Articulatory evidence for differentiating stress categories," in *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, edited by P. A. Keating (Cambridge U.P., Cambridge, England), Chap. 2, pp. 7–33.
 Beckman, M., Edwards, J., and Fletcher, J. (1992). "Prosodic structure and tempo in a sonority model of articulatory dynamics," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. Docherty and D. R. Ladd (Cambridge U.P., Cambridge, England), pp. 68–86.
 Beckman, M., and Pierrehumbert, J. (1986). "Intonational structure in English and Japanese," *Phonology* (Yearbook) 3, 255–310.
 Brownman, C., and Goldstein, L. (1992). "Articulatory phonology: an overview," *Phonetica* 49, 155–180.
 Byrd, D. (1996). "Influences on articulatory timing in consonant sequences," *J. Phon.* 24, 209–244.
 Byrd, D. M. (1994). "Articulatory timing in English consonant sequences," Ph.D. dissertation, UCLA, distributed as UCLA Working Papers in Phonetics 86, 1–196.
 Byrd, D., Flemming, E., Mueller, C. A., and Tan, C. C. (1995). "Using regions and indices in EPG data reduction," *J. Speech Hear. Res.* 38, 821–827.
 Byrd, D., Narayanan, S., Kaun, A., and Saltzman, E. (1997). "Phrasal signatures in articulation," to appear in *Proceedings of Laboratory Phonology V* (Cambridge U.P., Cambridge, England).
 Cole, R., Jakimik, J., and Cooper, W. (1978). "Perceptibility of phonetic features in fluent speech," *J. Acoust. Soc. Am.* 64, 44–56.
 Cooper, A. (1991). "Laryngeal and oral gestures in English /p,t,k/," *Proceedings of the XIIIth International Congress of Phonetic Sciences 2, University of Provence* (Aix-en-Provence, France), pp. 50–53.
 de Jong, K. (1995). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.* 97, 491–504.

Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (1996). "Glottalization of word-initial vowels as a function of prosodic structure," *J. Phonetics* 24, 423–444.
 Farnetani, E., and Vayra, M. (1996). "The role of prosody in the shaping of articulation in Italian CV syllables," in *Proceedings of the 1st ESCA Tutorial and Research Workshop on Speech Production Modeling and 4th Speech Production Seminar*, Autrans, France, edited by P. Perrier, F. Bussatret, and R. Laboissière (Institut National Polytechnique de Grenoble, France).
 Fougerson, C., and Keating, P. (1996). "Variations in velic and lingual articulation depending on prosodic position: Results for two French speakers," *UCLA Working Papers in Phonetics* 92, 88–96 (Also available electronically in the European Student Journal of Language and Speech at <http://web-sls.essex.ac.uk/web-sls/>).
 Fowler, C. A., and Saltzman, E. (1993). "Coordination and coarticulation in speech production," *Lang. Speech* 36, 171–195.
 Fromkin, V. (1965). "Some phonetic specifications of linguistic units: an electromyographic investigation," *UCLA Working Papers in Phonetics* 3, 1–184.
 Fujimura, O. (1990a). "Methods and goals of speech production research," *Lang. Speech* 33, 195–258.
 Fujimura, O. (1990b). "Articulatory perspectives of speech organization," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 323–342.
 Gelfer, C. E., Harris, K. S., and Baer, T. (1987). "Controlled variables in sentence intonation," in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, and K. S. Harris (Little, Brown, Boston, MA), pp. 422–435.
 Hawkins, J., and Cutler, A. (1988). "Psycholinguistic factors in morphological asymmetry," in *Explaining Language Universals*, edited by J. Hawkins (Blackwells, Oxford), pp. 281–317.
 Hinton, V. A. (1996). "Interlabial pressure during production of bilabial phones," *J. Phonetics* 24, 337–349.
 Hock, H. H. (1992). "Initial strengthening," in *Phonologica 1988*, edited by W. U. Dressler *et al.* (Cambridge U.P., Cambridge, England), pp. 101–110.
 Jun, S. (1993). "The phonetics and phonology of Korean prosody," Ph.D. dissertation, Ohio State University.
 Keating, P. A. (1995). "Effect of prosodic position on /t,d/ tongue/palate contact," poster presented at the XIIIth International Congress of Phonetic Sciences, Stockholm, 1995 (unpublished).
 Keating, P. A., Lindblom, B., Lubker, J., and Kreiman, J. (1994). "Variability in jaw height for segments in English and Swedish VCVs," *J. Phonetics* 22, 407–422.
 Klatt, D. H. (1975). "Vowel lengthening is syntactically determined in a connected discourse," *J. Phonetics* 3, 129–140.
 Krakow, R. A. (1989). "The Articulatory Organization of Syllables: A kinematic analysis of labial and velic gestures," Ph.D. dissertation, Yale University.
 Krakow, R. A., Bell-Berti, F., and Wang, Q. E. (1994). "Supralaryngeal declination: evidence from the velum," in *Producing Speech: A Festschrift for Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (AIP, Woodbury, NY), Chap. 23, pp. 333–353.
 Ladd, D. R. (1984). "Declination: A review and some hypotheses," *Phonology* (Yearbook) 1, 53–74.
 Ladefoged, P. (1967). *Three Areas of Experimental Phonetics* (Oxford U.P., London).
 Larkey, L. (1983). "Reiterant speech: An acoustic and perceptual validation," *J. Acoust. Soc. Am.* 73, 1337–1345.
 Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* 35, 1773–1781.
 Löfqvist, A. (1996). Personal communication.
 Maeda, S. (1976). "A characterization of American English Intonation," Ph.D. dissertation, MIT.
 Moon, S. J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* 96, 40–55.
 Öhman, S. (1967). "Word and sentence intonation: a quantitative model," *STLQ Progr. Status Rep.* Stockholm 2/3, 20–54.
 Oller, D. K. (1973). "The effect of position in utterance on speech segment duration in English," *J. Acoust. Soc. Am.* 54, 1235–1247.
 Pierrehumbert, J., and Beckman, M. (1988). *Japanese Tone Structure* (MIT, Cambridge, MA).
 Pierrehumbert, J., and Talkin, D. (1992). "Lenition of /h/ and glottal stop,"

- in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. Docherty and D. R. Ladd (Cambridge U.P., Cambridge, England), Chap. 4, pp. 90–117.
- Shattuck-Hufnagel, S., and Turk, A. (1996). “A prosody tutorial for investigators of auditory sentence processing,” *J. Psycholinguistic Res.* **25**(2), 193–247.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). “TOBI: A Standard for Labeling English Prosody,” *1992 International Conference on Spoken Language Processing*, edited by J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, and G. E. Wiebe (University of Alberta, Edmonton, Canada), Vol. 2, pp. 867–870.
- Sluijter, A. M. C., van Heuven, V. J., and Pacilly, J. J. A. (1997). “Spectral balance as a cue in the perception of linguistic stress,” *J. Acoust. Soc. Am.* **101**, 503–513.
- Straka, G. (1963). “La division des sons du langage en voyelles et consonnes: peut-elle être justifiée?” *Travaux de Linguistique et de Littérature*, Centre de Philologie et de Littératures Romanes de l’Université de Strasbourg, Vol. I, pp. 17–99.
- ’t Hart, J., Collier, R., and Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*, Cambridge Studies in Speech Science and Communication (Cambridge U.P., Cambridge, England), Chap. 5, pp. 121–150.
- Thorsen, N. (1985). “Intonation and text in Standard Danish,” *J. Acoust. Soc. Am.* **77**, 1205–1216.
- Vaissière, J. (1986). “Comment on Abbs’s Paper,” in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 220–222.
- Vaissière, J. (1988). “Prediction of velum movement from phonological specifications,” *Phonetica* **45**, 122–139.
- van Lieshout, P. H., Starkweather, C. W., Hulstijn, W., and Peters, H. F. M. (1995). “Effects of linguistic correlates of stuttering on emg activity in nonstuttering speakers,” *J. Speech Hear. Res.* **38**, 360–372.
- Vayra, M., and Fowler, C. (1992). “Declination of supralaryngeal gestures in spoken Italian,” *Phonetica* **49**, 48–60.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). “Segmental durations in the vicinity of prosodic phrase boundaries,” *J. Acoust. Soc. Am.* **92**, 1707–17.

Coarticulatory stability in American English /r/

Suzanne Boyce and Carol Y. Espy-Wilson

Electrical, Computer, and Systems Engineering Department, Boston University, 44 Cummington Street, Boston, Massachusetts 02215 and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 15 May 1995; accepted for publication 4 February 1997)

A number of different researchers have reported a substantial degree of variability in how American English /r/ coarticulates with neighboring segments. Acoustic and articulatory data were used to investigate this variability for speakers of “rhotic” American English dialects. Three issues were addressed: (1) the degree to which the $F3$ trajectory is affected by segmental context and stress, (2) to what extent the data support a “coproduction” versus a “spreading” model of coarticulation, and (3) the degree to which the major acoustic manifestation of American English /r/—the time course of $F3$ —reflects tongue movement for /r/. The $F3$ formant trajectory durations were measured by automatic procedure and compared for nonsense words of the form /waCrav/ and /wa'Crav/, where C indicates a labial, alveolar, or velar consonant. These durations were compared to $F3$ trajectory durations in /warav/ and /wa'rav/. In addition, formant values in initial syllables of words with and without /r/ were examined for effects of intervening consonant contexts. Results indicated similar $F3$ trajectory durations across the different consonant contexts, and to a lesser degree across stress, suggesting that coarticulation of /r/ can be achieved by overlap of a stable /r/-related articulatory trajectory with movements for neighboring sounds. This interpretation, and the concordance of $F3$ time course with tongue movement for /r/, was supported by direct measures of tongue movement for one subject. © 1997 Acoustical Society of America. [S0001-4966(97)02106-1]

PACS numbers: 43.70.Bk, 43.70.Fq [AL]

INTRODUCTION

In the standard “rhotic” dialects of American English (where /r/ is pronounced in all allowable contexts), /r/ has been described as coarticulating with adjacent segments in a number of interesting ways. The best known of these effects involve vowels. For instance, vowels next to consonantal /r/ show coarticulatory effects known as “/r/-coloring” (Ladefoged, 1982; Giergerich, 1992; Bronstein, 1967). However, coarticulatory effects on neighboring consonants have also been described (Olive *et al.*, 1993; Shoup and Pfeifer, 1976; Zue, 1985). For speech recognition systems, this variability can result in the misclassification of nearby vowels and/or consonants as /r/ (Espy-Wilson, 1994).

This paper is concerned with acoustic and articulatory aspects of the way consonantal /r/ interacts with adjoining consonant and vowel segments in “rhotic” varieties of American English. Because /r/ as produced by American English speakers appears to involve several articulators acting in concert and shows wide variability in articulatory configuration between speakers, we concentrate on analysis of consistency in its acoustic signature. The results we describe are important for phonological descriptions of American English, and for the design of speech recognition systems, as well as for models of motor control in normal and disordered speech.

A. Acoustics of /r/

The most salient feature of American English /r/, whether consonantal or vocalic, is its low $F3$, which can range between 1100 and 2000 Hz but which is normally in

the region of 1600 Hz for both men and women (Espy-Wilson, 1992, 1994; Nolan, 1983; Lehiste and Peterson, 1961; Lehiste, 1962; Zue, 1985; but cf. Hagiwara, 1995 for an examination of male-female differences). For other segments of American English, the typical $F3$ range occurs between 2100 and 3000 Hz (Peterson and Barney, 1952; Shoup and Pfeifer, 1976). Typically, the $F3$ transition between surrounding segments and /r/ shows a marked trajectory of movement beginning at 2000 Hz. When /r/ is surrounded by sonorant segments, a complete $F3$ trajectory representing movement toward and away from the articulatory configuration for /r/ can be seen. For all types of /r/ this trajectory resembles an inverted parabola. In general, it is reasonable to assume that the time course of frequency change in $F3$ below 2000 Hz reflects the time course of articulatory movement specific to /r/. In other words, a parabolic $F3$ trajectory below 2000 Hz reflects /r/-related movement.¹

In a study of semivowels, Espy-Wilson (1992, 1994) found that when lowering to 2000 Hz or below was used as a criterion for /r/ identification in a speech recognition system, segments adjacent to /r/ were routinely misidentified as /r/ proper. This result reflected the fact that $F3$ values were frequently lowest on these adjacent segments, while $F3$ values during the segment transcribed as /r/ were somewhat higher. A typical example is illustrated in Fig. 1, which shows a spectrogram of the word “everyday” spoken by a native American English speaker. The results of a formant-tracking program (Espy-Wilson, 1987) have been superimposed on the spectrogram. Vertical lines in the phonetic transcription at top show the boundaries of /r/ and neighboring segments as assigned by a standard acoustic segmentation

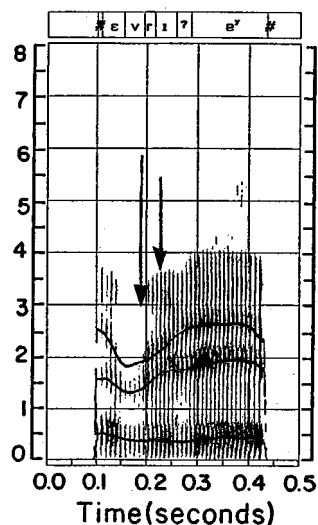


FIG. 1. Spectrogram of the word “everyday” with formant tracks overlaid and phonetic transcription at top. Arrows indicate boundaries assigned to the /r/ by the segmentation procedure. The lowest point of F_3 occurs outside of the boundaries assigned to /r/ by segmentation procedure.

procedure (Seneff and Zue, 1988).² Note that this word shows a full, parabolic F_3 trajectory typical of /r/. Two other facts stand out: (1) the lowest point of F_3 (which we noted above probably corresponds to the most extreme /r/-related movement) occurs during the preceding fricative, and (2) inside the boundaries assigned to /r/ proper we see the rising portion of the acoustic /r/ trajectory (corresponding to offset from the articulatory extremum). Thus “variability” in the instantiation of /r/ for this word appears to involve variability in the way the articulatory movement (and associated acoustic parabolic trajectory) is placed with respect to segmental boundaries. Additionally, the F_3 trajectory appears continuous through articulation of the labial obstruent.

B. Models of coarticulation

Classically, coarticulation is defined as an assimilation in the articulation of one segment, a “target” segment, as a result of a neighboring “home” segment. It is said to occur when the effects of one segment show up during production of another segment. Physically, coarticulation may be manifested as a change in dynamic characteristics of movement (shape/displacement/duration of the articulatory movement) as well as change in placement within the vocal tract. Because articulatory postures are attained dynamically, through movements whose trajectory exhibits a defined onset, extremum, and offset, trajectory duration may increase as a result of a longer onset, a plateau of movement at the extremum, or a longer offset. Shape may change as a result of durational change in any of these components, or because of change in the displacement of the movement.

Current theories of anticipatory coarticulation, i.e., coarticulation between a target and a following home segment, explain these effects in one of two ways. In one approach, known as the “feature-spreading” or “spreading” account, the underlying articulatory plan (including trajectory of movement, placement of movement in the vocal tract, etc.) for producing the target segment has been altered from the

form it would take if it were bordered by different neighboring segments; in other words, articulatory plan varies by segmental context (Daniloff and Moll, 1968; Hammarberg, 1976; Kent and Minifie, 1977; Keating, 1988; also see Perkell and Matthies, 1992, among others). Articulatorily, this conception has two critical assumptions: (1) coarticulatory effects occur because articulatory postures associated with the “home” segment are achieved over an extended period of time (longer than required for the home segment *per se*), and (2) the degree of coarticulatory change will vary according to the difficulty of sustaining simultaneously the “home” and “target” articulatory postures. In particular, it is assumed that if the “home” and “target” specifications are easy to reconcile, the two segments will be coarticulated for a longer period of time, while if the “home” and “target” are difficult to reconcile (an articulator directed to be in two places at once, for instance), there will be less coarticulation. For example, lip retraction required for the vowel /i/ is considered to conflict with coarticulatory spreading of lip rounding (Hammarberg, 1976; Perkell and Matthies, 1992). In contrast, anticipatory spreading of coarticulation is expected to be at a maximum when adjacent segments involve different articulators. For instance, anticipation of tongue movement for vowels such as /i/ or /a/ is considered to be maximal when the preceding consonant is a labial, since the tongue is theoretically free to move (Harris and Bell-Berti, 1984). Moreover, unrounded consonants such as /s/ and /t/ are often assumed to be potentially compatible with rounding coarticulation (cf. Perkell and Matthies, 1992; Boyce *et al.*, 1990 for overview). Predictions of coarticulatory effect are less clear when adjacent segments require movement by the same articulator in similar but not identical directions or to similar but not identical positions in the vocal tract. Investigators looking at cases such as the interaction of tongue dorsum movement for /u/ or /i/ and adjacent /k/, have concluded that the (observed) articulatory trajectories of both tend to be affected, according to constraints on individual segments (Recasens, 1985). A schematic illustration of one version of the spreading model, showing the contrast between movement for isolated segments versus segments in context, is illustrated in Fig. 2(a) and (b).

A different view, known as the “coproduction” approach, is that much of what we call coarticulation can be explained, not by changing the segmental articulatory plan, but as the result of overlap and consequent “blending” between (unaltered) articulatory plans for adjacent segments; i.e., specified articulatory trajectories for adjacent target and home segments combine to produce a movement trajectory that is intermediate between them (Munhall and Lofqvist, 1992; Gracco and Lofqvist, 1994). Arguments using articulatory and acoustic data to support this view have been advanced by Harris and Bell-Berti (1984), Gelfer *et al.* (1989), Boyce *et al.* (1990), Browman and Goldstein (1986), Browman and Goldstein (1990), Bell-Berti and Krakow (1991), Fowler (1993), and Bell-Berti *et al.* (1995) among others. This model is schematized in Fig. 2(c). Some critical assumptions of this view include (1) the underlying motor programming for articulatory movement to and from the extremum articulatory configuration remains relatively stable

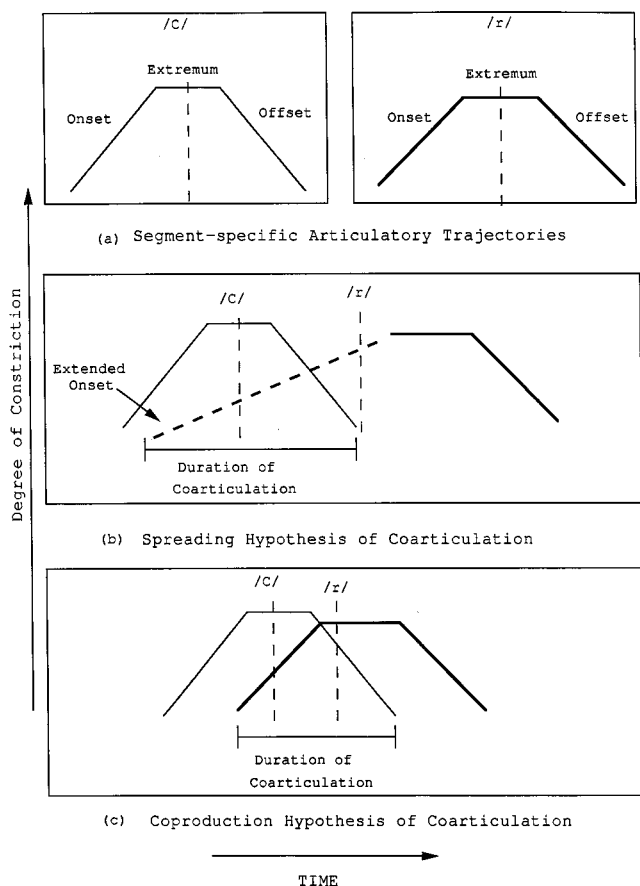


FIG. 2. A schematization of two viewpoints of coarticulation: (a) individual gestures for a consonant and /r/, (b) coarticulation based on one version of the spreading model (Perkell and Matthies, 1992), (c) coarticulation based on the coproduction model.

across segmental contexts, and (2) underlying trajectories must be deduced from observed trajectories by judicious examination of observed trajectories across different segmental contexts. An important aspect of this view is that the speech motor system prefers to maintain segmental plans, and stable trajectories, when possible. Proponents of the coproduction viewpoint have also suggested that accommodation to the particular requirements of segmental context may be accomplished by displacement, or “sliding” of articulatory movement trajectories away from their home segment; in other words, the extremum of the articulatory posture, and thus its spatiotemporally stable onset, extremum and offset, can be shifted in time (Browman and Goldstein, 1986, 1990). For instance, difficult interactions between specifications on adjacent segments may be mediated by changing the spacing of associated articulatory movements in time. Changes in speech rate, stress, and syllable position, etc. may be accomplished either by changes in the segmental articulatory plan (Gracco and Lofqvist, 1994) or by “sliding” (Browman and Goldstein, 1986, 1990).

In general, data suggesting changes in articulatory trajectories due to context (and not attributable to blending) would constitute support for the “spreading” approach, while data suggesting stable trajectories would constitute support for the “coproduction” approach. For /r/, it is reasonable to assume that changes in the acoustic trajectory spe-

cific to /r/ will reflect changes in the articulatory instantiation of /r/. Thus, the spreading model predicts that we will see changes in the duration and/or shape of acoustic /r/ trajectories across segmental context. In contrast, the “coproduction” model predicts stability in the /r/-related $F3$ trajectory.

This issue is particularly important for /r/, because the use of “retroflex,” “bunched,” and “mixed” versions of /r/ (which use different combinations of tongue tip and tongue dorsum to make constrictions along the palate) varies in a nonobvious way among the population (Delattre, 1967; Delattre and Freeman, 1968; Bernthal and Bankson, 1993; Lindau, 1985; Hagiwara, 1995; Westbury *et al.*, 1995; Narayanan *et al.*, 1997). Further, complex interactions between pharyngeal constriction, labial constriction, and different types of tongue constrictions during /r/ makes coarticulatory conflicts hard to predict. Given these caveats, however, we can make the following generalizations. (1) For all /r/’s, we might expect coarticulation to occur most freely (and in the spreading model, trajectory duration/shape to change the most) when adjacent segments do not involve the tongue at all, i.e., labial and glottal consonants. (2) In addition, we might expect that the /r/ variant used by the subject would affect the way /r/ coarticulates with surrounding segments. For instance, a subject who uses his tongue tip primarily to make an oral cavity constriction for /r/ in a vocalic context might show different contextual effects on the /r/ trajectory when neighboring segments involve the tongue tip versus the tongue dorsum. A similar argument can be made for subjects who use primarily the tongue dorsum during /r/. Speakers may also respond to the difficulty of sequencing /r/ with alveolar or velar consonants by alternating between /r/ variants according to context (Espy-Wilson and Boyce, 1994). Each of these possibilities (and others not mentioned) suggests different scenarios, depending on the particular characteristics of the /r/ variant used, and the particular articulatory interactions involved. Thus although much remains unknown about how /r/ coarticulates with surrounding phones, it seems reasonable to assume that (1) for purposes of the spreading hypothesis, labial contexts provide fewer challenges to coarticulation than velar or alveolar contexts, and (2) if context has any effect, we might expect this to emerge in a comparison of the shape and duration of the $F3$ trajectory for /r/ across vocalic, labial, velar, and alveolar contexts.

I. GENERAL METHODOLOGY

Data for this study include acoustic signal data recorded from seven speakers (three female and four male) and articulatory tongue movement data recorded from one of the male speakers. Articulatory data were used to confirm methodology and conclusions from acoustic data. Methodology and results specific to the acoustic data are described in experiment 1. Methodology and results specific to articulatory data are described in experiment 2. Methodology shared between experiments is described below.

Seven speakers produced five repetitions of experimental nonsense words /wavrav/, /wabrav/, /wagrav/, /wadrav/, and five repetitions of the control nonsense words /warav/, /wawav/, /wavav/, /wabav/, /wagav/, and /wadav/. Each

nonsense word was produced in two stress conditions: with stress on the first syllable (initial stress) and with stress on the second syllable (final stress); e.g., /^ˈwarav/, /wa^ˈrav/. As a control for nonsense word effects, three speakers produced five repetitions each of a smaller set of real words structured to resemble a representative sample of the nonsense words, e.g., “Africa,” “begrime,” and “barometer” plus cases of word-initial and word-final /r/ such as “rob” and “bar.” All words were embedded in the carrier phrase “Say ——— for me.”

The subjects produced the experimental stimuli in the same order five times with reference to a handheld paper list. For all subjects except RD, acoustic signals were digitized at 16 kHz on a SUN workstation via the ESPS/WAVES signal processing software. For subject RD, the acoustic signal was digitized at 10 kHz on a DEC workstation via the MITSYN signal processing software. For analysis, the signals were transported to a SUN workstation and subjected to signal processing using ESPS/WAVES.

Subjects were speakers of fully rhotic versions of standard American English from Missouri, western Massachusetts, upper New York state, western Pennsylvania, Michigan, Philadelphia, and Washington state. Speakers were instructed to produce words at a self-selected comfortable and consistent rate, in a natural manner, and were given a short practice session. Of the seven subjects, four (three females and one male) were phonetically sophisticated, three (all males) were not. At the time of recording, speakers HSS, BS, and MS had some notion of the purpose of the study; the four male speakers HD and RD had none.³ The experimental nonsense words were designed to include cases with labial, alveolar, and velar consonants before /r/. The control words were included to allow analysis of the formant trajectories characteristic of these consonants as well as those of the labial most like /r/ (/w/) and of /r/ itself. Additionally, the comparison between /g/ and /w/ provided a rough indication of the extent of *F3* lowering attributable to rounding. Segments following /r/ in the experimental nonsense words were the same across words, allowing consistent comparison in the raising portion of the *F3* trajectories. The experimental words /wabrav/ and /wavrav/ were expected to present the most favorable conditions for coarticulation of /r/; that is, we expected that if “spreading” of /r/ articulation occurs, these words would show longer *F3* trajectories (and presumably longer articulatory trajectories) than those for words with singleton /r/. If articulatory movements for a segment are spatiotemporally stable, as predicted by the coproduction model, then we expected trajectories to be the same as those for words with singleton /r/. The words /wagrav/ and /wadrav/, because /g/ and /d/ involve the tongue, were expected to represent more difficult coarticulatory challenges. The spreading model predicts that /r/ trajectories in such words would be shorter than in control words such as /warav/, or words with labial consonants such as /wavrav/.

II. EXPERIMENT 1: ACOUSTIC INVESTIGATIONS

Acoustic data were used to determine (a) whether stress and consonantal context affect *F3* trajectory durations, and (b) if changes in stress, context, and/or trajectory duration may affect the shape of *F3* trajectories.

A. Methodology

All subjects except RD were recorded in a sound-treated room using a Sennheiser directional microphone and a high quality Yamaha audio cassette tape recorder. Subject RD was recorded in a quiet hard-walled room using a high-quality SONY directional electret condenser microphone. The acoustic signal for RD was recorded digitally on line using the MITSYN signal processing software.

Formant tracks were computed for all the utterances using the ESPS/WAVES formant tracker and a 10-ms frame rate. Alignment between formant tracks and spectrograms was handled automatically as part of the WAVES program. For the purpose of analyzing *F3* trajectory duration and shape, the formant tracks were edited by the two authors working together to eliminate noisy or erroneous data points as described below. To cut down on editing, for each word the three tokens best analyzed by the WAVES formant tracker were chosen (in some instances, more tokens were included). Files containing *F3* values from edited formant tracks were transferred to a Macintosh IIsi computer and analyzed using standard graphics and statistics programs.

All editing was done by visual reference to spectrograms for each token with results of the formant tracker superimposed, and power spectra where appropriate. Figure 3 shows illustrative spectrograms with superimposed formant tracks for tokens /warav/, /wavrav/, /wadrav/, and /wagrav/ produced by speaker JM. Several steps were involved in editing the formant tracks. First, *F3* tracks during the word-initial /w/ were deleted. The criterion for the start of the following vowel (V1) was the beginning of strong energy in *F1*. Second, *F3* tracks during the word-final /v/ were deleted. The criterion for the end of the second vowel (V2) was the end of strong energy in *F1*. If the formant tracks appeared continuous and unambiguous, as in panel (a) of Fig. 3, no further editing was done. If the *F3* tracks during the intervocalic obstruents were noisy, as in panel (c) of Fig. 3, the frequency values were deleted while maintaining the correct spacing in time between retained values. The criterion for deletion of frequency values was *F3* spectral amplitude 30 dB or more below the amplitude at the lower frequency spectral peak. Parts (b) of the two panels of Fig. 4 show examples of edited formant tracks with and without deletion of noisy values. Note that because the initial and final consonants are eliminated from the edited version, the edited versions are shorter than the spectrographic version.

In some cases obstruents were produced with incomplete vocal tract closure, and the formant tracker was able to detect consistent and appropriate *F3* values in at least some portion of the acoustically defined closure interval. These values were retained under any one of the following conditions: (a) there was little or no stop burst, (b) energy at low frequencies was present throughout the closure interval, and (c) the time course of the formant tracks was similar over the five repeti-

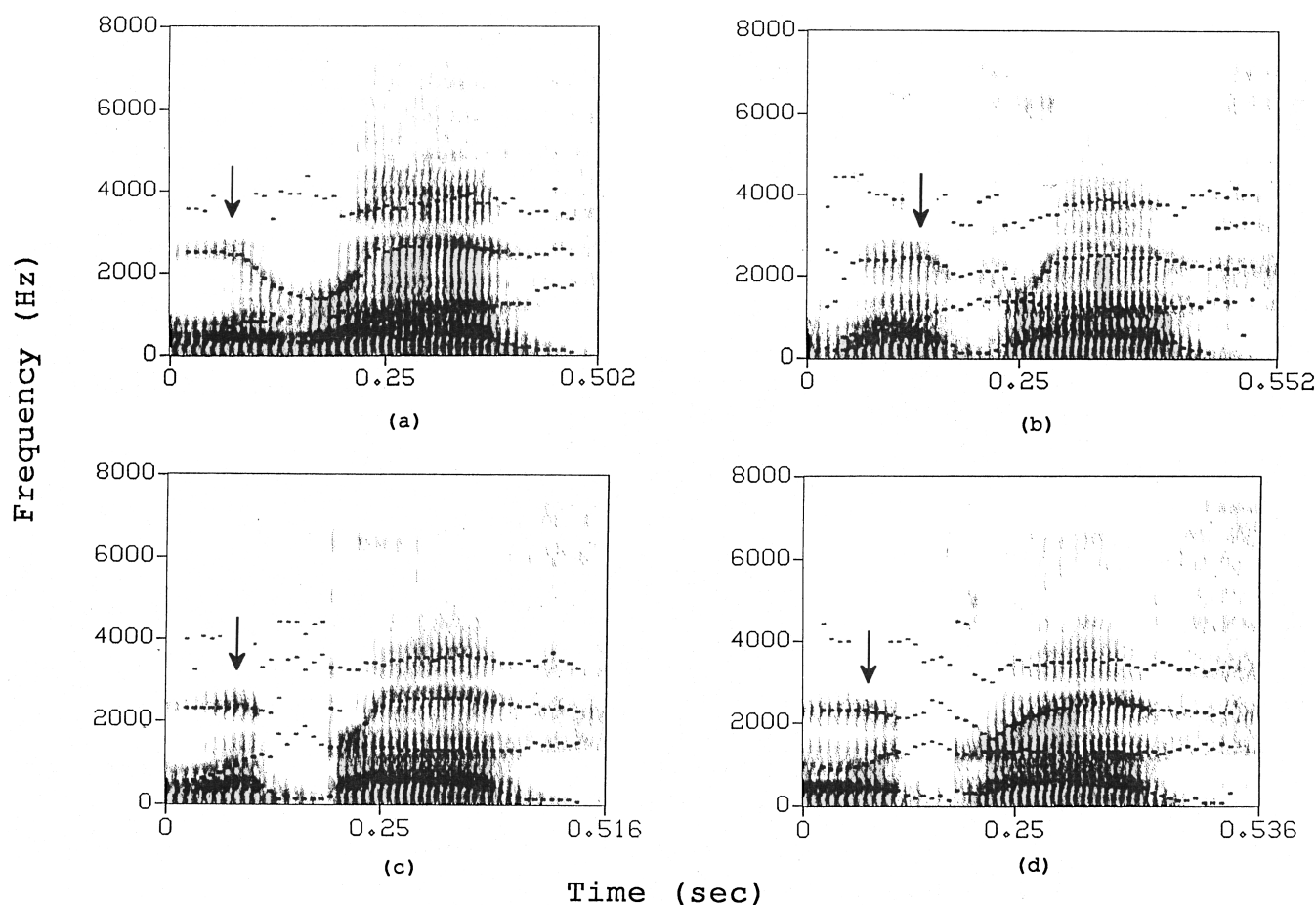


FIG. 3. Spectrograms of (a) /wa'rarv/, (b) /wa'rvav/, (c) /wa'brav/, and (d) /wa'grav/ produced by speaker JM. Arrows point to syllabic peaks found during the first syllable by an automatic procedure.

tions produced by each speaker. A relatively unambiguous example of this type can be found in the left panel of Fig. 4. In some cases, the formant tracker incorrectly assigned values belonging to F_3 as belonging to F_2 or F_4 ; these values were replaced by the correct values. Additionally, there were ambiguous cases in which the formant tracker identified energy simultaneously at two points in the spectrum which might plausibly reflect F_3 . Examples of these cases are shown in panels (b) and (d) of Fig. 3. (Generally, the two paths would be represented as belonging to F_2 and F_3 , or F_3 and F_4 , or some mixture of the two.) Almost invariably in these cases of “double” paths, one track resembled the pattern of F_3 seen during the closure in control words /wadv/, /wagav/, /wabav/, /wavav/, while the visible portion of the other track resembled the pattern seen for /r/ in /warav/. Our strategy for dealing with these cases is described below (see Sec. II A 1). In uncertain cases, formant values were determined to be valid or invalid by referring to formant patterns in the control utterances, including /warav/.

1. “Double” F_3 paths

Cases in the data with discernible “double path” resonances (i.e., there were two simultaneous resonances that might be called F_3) occurred in all consonant contexts, for both stress conditions, and for all speakers. Parallel cases of “double” trajectories for representative tokens of /'wavrav/

(produced by speaker RD), and /wa'drav/ (produced by speaker JM), are shown in Figs. 5 and 6. These are contrasted with representative tokens of /'warav/ or /wa'rarv/, and control words /'wavav/ or /wa'dav/, as appropriate. In the /'wavrav/ case, energy was present throughout the fricative constriction, and formant tracks for both paths were relatively continuous. In the /wa'drav/ case, the formant tracks show evidence of both paths, but in a less continuous fashion.

The pattern shown by resonances in the range 1500–2500 Hz during the intervocalic intervals of /'wavrav/ (Fig. 5) is typical. At the end of the initial vowel (V1) two resonances appear that might be labeled F_3 ; these are greatly attenuated during consonant constriction (depending on the degree of constriction) but may still be discerned in the signal. As can be seen, the “lower path” resonance trajectory in /'wavrav/ is quite similar to what we see for the F_3 trajectory in the control word /'warav/. However, if we follow the “upper” path resonance trajectory, two points stand out: (1) that the falling-rising portion of the trajectory is much shorter than that seen in /'warav/ and occurs a considerable time after the end of V1, and (2) that immediately after the end of V1, and during the /v/ constriction, the “upper path” resonance trajectory resembles the F_3 values tracked during the /v/ constriction in /'wavrav/. The “double” resonance pattern of the /wa'drav/ token shown in Fig. 6 bears a simi-

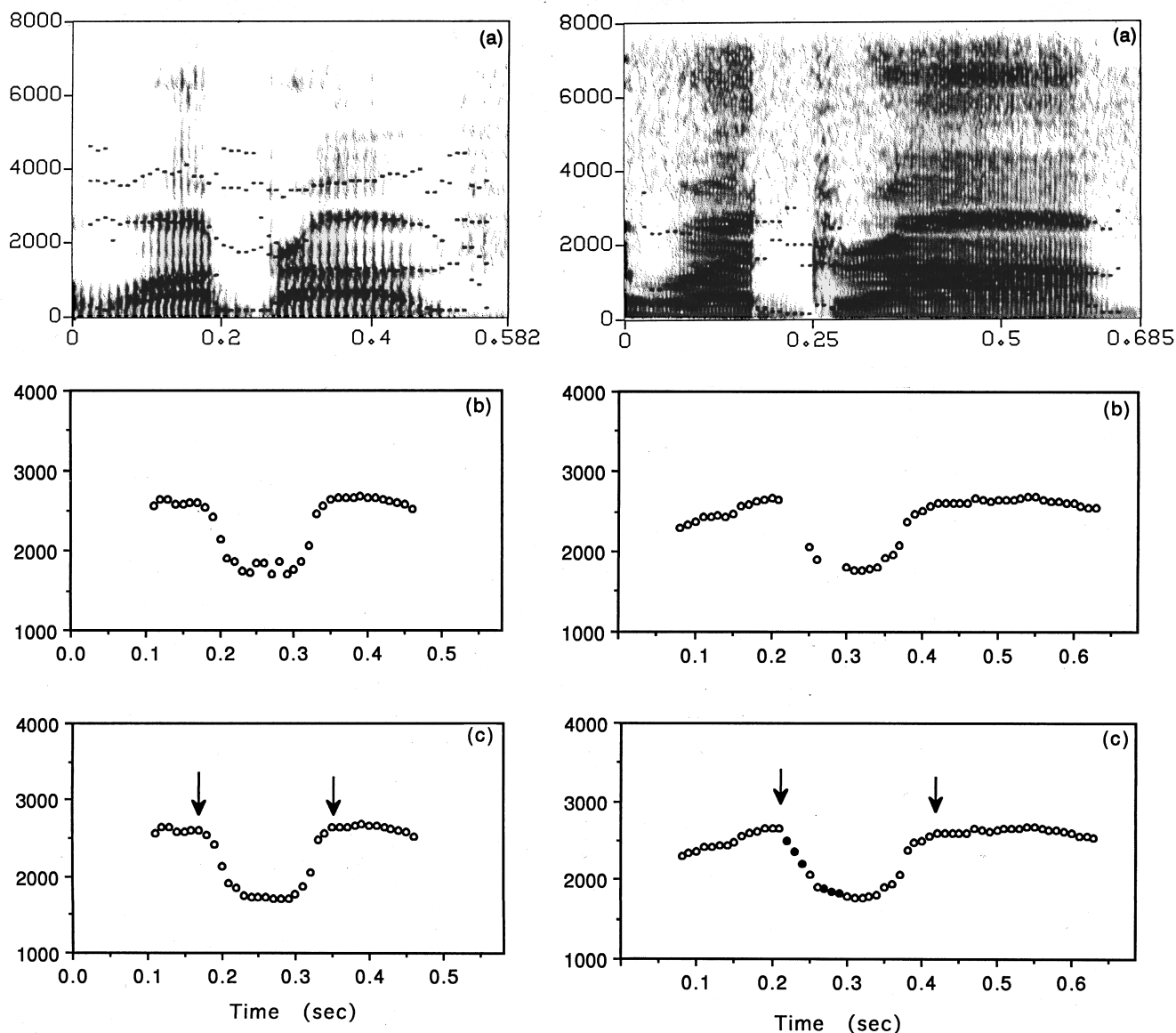


FIG. 4. (a) Spectrograms of /wa'brav/ produced by JM (left panel) and /wa'drav/ produced by HSS (right panel) with formant tracks overlaid. (b) Edited F_3 tracks. (c) Smoothed and/or interpolated edited F_3 tracks with automatically determined inflection points indicated with arrows.

lar relationship to the F_3 trajectories of /wa'rav/ and /wa'dav/. In this case, however, while the initial lowering of F_3 at the end of the vowel is evident, there are missing values during the /d/ closure. As with the /'wavrav/ case discussed above, the observable portion of the lower path aligns well with the F_3 trajectory visible in /'warav/, whereas the "upper" path trajectory resembles that of the /d/ in /'wadav/. The shorter duration of the fall-rise portion of the "upper path" trajectory can be attributed to constriction narrowing for the contextual consonant rather than the slower articulation of /r/. Altogether, it seems clear that the "upper path" trajectory in these situations reflects the influence of the obstruent preceding /r/, while the "lower path" trajectory reflects the influence of the /r/. (Presumably, the variation we see in whether "double" resonances can be discerned in the signal, and whether the "upper path," or "lower path" resonance is stronger, can be attributed to normal token-to-token variation in the articulation of consonant

and /r/ segments.) This reasoning was confirmed by articulatory data from RD (see Sec. III). Thus, the "lower path" F_3 trajectory was the object of measurement in all cases.

2. Identifying trajectory end points

Because visual identification of trajectory beginning and end would be subject to experimenter bias, an automatic procedure was developed to identify trajectory beginning and end points for the /r/-related F_3 trajectory of initial- and final-stress tokens of the /warav/, /wabrav/, /wavrav/, /wadrav/, and /wagrav/ nonsense words. Typically, trajectories in these data show some gradual lowering and raising movement on the periphery prior to, and following, an identifiable "bend" associated with /r/. We defined trajectory edges as inflection points at these "bends," and duration of the trajectory as the time between inflection points. Our program found these inflection points based on a combination of the first and second differences of the F_3 trajectory. When

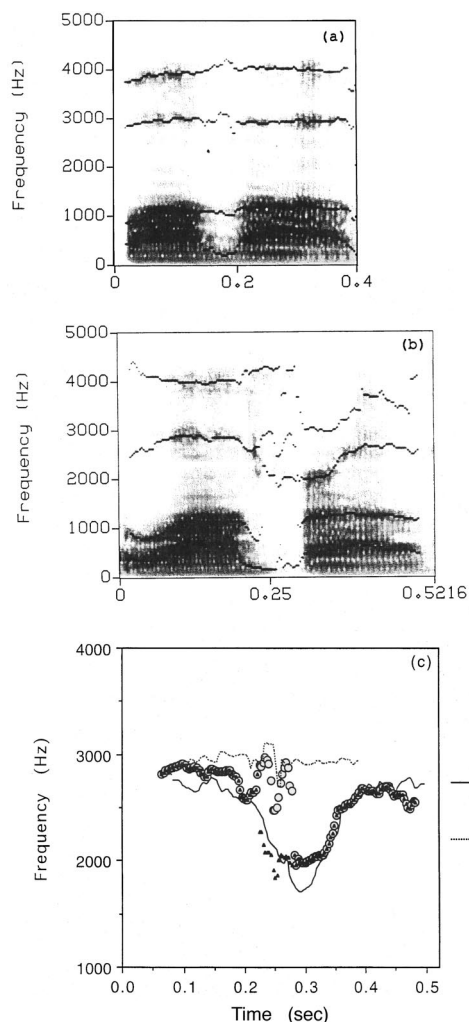


FIG. 5. (a) Spectrogram of /wavy/ with formant tracks overlaid. (b) Spectrogram of /wavy/ with formant tracks overlaid that show two paths for F_3 during the /v/. (c) Comparison of formant tracks taken from one token of /wavy/ and from /wavy/ and /wavy/. Data are from speaker RD. Note that part (c) is repeated in Fig. 9 which shows alignment with articulatory data.

two competing inflection points were found, the more peripheral was used. Minor local perturbations in the F_3 tracks were smoothed by hand, and missing values (corresponding to ill-tracked or noisy values eliminated during editing) were filled in by a simple linear interpolation algorithm, producing a continuous trajectory. Parts (c) of Fig. 4 show examples of interpolated F_3 tracks where the interpolated values are indicated by filled squares. Arrows show inflection points as found by this algorithm.

Variability due to the automatic procedure was of two types. First, because the automatic procedure was forbidden to assign trajectory beginning during the interpolated portion, and trajectory beginning was typically identified on the left edge of the interpolated region, the automatic procedure tended to find slightly longer trajectory durations for /Cr/ word tokens with noisy consonant closure intervals, in contrast to measurements for tokens with voicing through the consonant closure interval or for singleton /r/ words. Second, minor differences in trajectory slope on the right and left edges could affect the determination of inflection point.

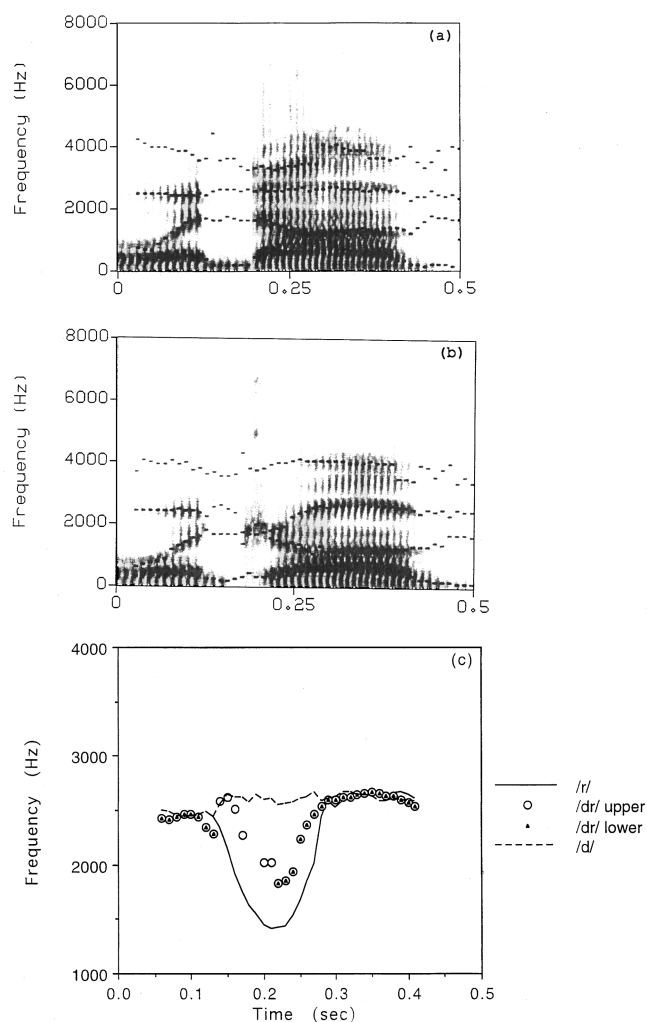


FIG. 6. (a) Spectrogram of /wavy/ with formants tracks overlaid. (b) Spectrogram of /wavy/ with formants tracks overlaid that show two paths for F_3 during the /d/ closure. (c) Comparison of formant tracks taken from one token of /wavy/ [the spectrogram of this token is shown in Fig. 3(a)] and from /wavy/ and /wavy/. Data are from speaker JM.

(Slope was in turn dependent on the F_3 values for initial and final vowels, which were idiosyncratic to speaker as well as to degree of vowel reduction and stress.) An error of 1 sample point in locating trajectory end points corresponded to an error of 10 ms, due to the 10-ms frame rate for the formant tracker. Tokens where the automatic procedure missed a visually identifiable peripheral inflection point were adjusted by hand. We estimate error conservatively at ± 20 ms for each trajectory end point. Thus two trajectories of equal duration might conceivably be measured as different by 40 ms. Figure 7 illustrates the typical situation found across tokens for all subjects in our study. Although this token of /wavy/ and two tokens of /wavy/ produced by speaker HD have extremely similar F_3 trajectories, and visual measurement would identify very similar trajectory beginning and ending points, sensitivity to minor differences in slope caused the automatic procedure to calculate the duration differences between inflection points (i.e., the trajectory durations) as 180 ms for the singleton /r/ token (inflection points indicated by open arrows) versus 220 and 230 ms for the two /vr/ tokens (inflection points indicated by filled ar-

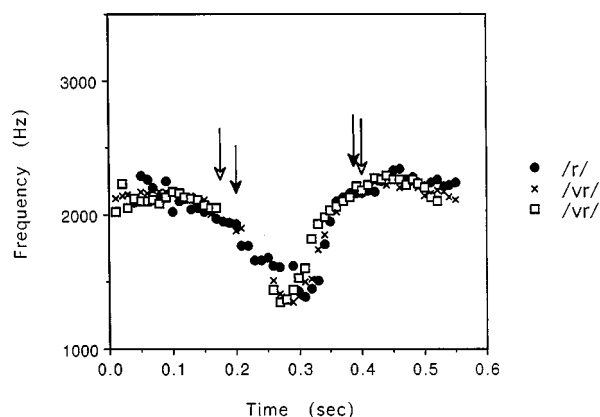


FIG. 7. Edited F_3 tracks of one token of /wa'rav/ and two tokens of /wa'vrv/ produced by HD. Arrows show automatically assigned inflection points.

rows). The token of /wa'grav/ illustrated in Fig. 8 is another case in point; although the similarity between /wa'grav/ and other tokens is patent, a slightly more gradual slope along the right-hand edge caused the /wa'grav/ token trajectory duration to be computed as 250 ms. (Other tokens of /wa'grav/ for this subject showed similar trajectories but were measured at 210 and 190 ms.) Similarly, the durations of the remaining trajectories in Fig. 8 vary between 160 ms (in the case of /wa'rav/, inflection points indicated by filled arrows) and 200 ms (in the case of /wa'brav/, inflection points indicated by open arrows). Random pairings of tokens across the dataset for all speakers showed parallel patterns of variability. Thus, our measurement procedure appeared likely to overestimate the true variability of the dataset. Although this was not ideal, any findings of consistent behavior were unlikely to be artifactual in nature.

B. Qualitative and quantitative results

The feature-spreading account of coarticulation predicts that the F_3 trajectory for /r/ will vary in trajectory shape and duration across contexts (although its visibility may be obscured at points by token-to-token variation and by the

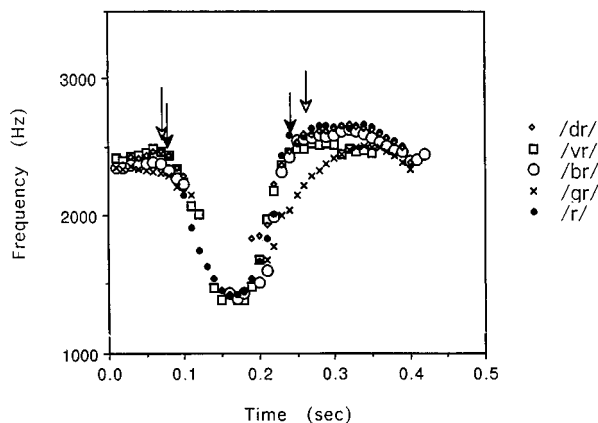


FIG. 8. Edited F_3 tracks of one token of /wa'rav/, /wa'vrv/, /wa'brav/, /wa'drav/, and /wa'grav/ produced by JM (smoothing and interpolation are not shown). Spectrograms of these tokens are shown in Fig. 3 and part (b) of Fig. 6.

acoustic effects of neighboring sounds). In contrast, the data showed consistent evidence of trajectory similarity across the dataset. This similarity was maintained between tokens of the same word, across different consonant contexts, and to a large degree across stress.

1. Shape of F_3 trajectory

Figure 7 shows edited F_3 tracks (before interpolation or smoothing) extracted from one token of /wa'rav/ and two tokens of /wa'vrv/ spoken by subject HD. All tokens are lined up at the beginning of V1, which was of approximately the same duration for each. It is clear that the F_3 trajectories of the /wa'vrv/ tokens show extremely similar falling–rising shape and duration, as predicted. [To emphasize the similarity between these trajectories and that for /wa'rav/ (which does not include a consonant interval), the F_3 trajectory for /r/ in /wa'rav/ was shifted to the right by 30 ms.] This token-to-token similarity in F_3 trajectory shape was consistent across the dataset, although for some tokens the trajectories might begin earlier or later in time. A similar picture for tokens of different consonant-/r/ combinations and singleton /r/ can be seen in Fig. 8, which shows the edited F_3 tracks taken from one token each of /wa'vrv/, /wa'brav/, /wa'grav/, /wa'drav/, and /wa'rav/ for the same subject. The tokens of /wa'vrv/, /wa'grav/, and /wa'brav/ are lined up at the beginning of V1. For all tokens except /wa'drav/, the duration of V1 and of the occlusion interval were approximately the same. (Here the F_3 trajectory for /r/ in /wa'rav/ was shifted to the right by 70 ms.) Because for this token of /wa'drav/ the vowel and occlusion interval were slightly shorter than those for the other tokens, the /wa'drav/ trajectory was shifted to the right by 20 ms. Spectrograms (with formant tracks superimposed) of these words are shown in Figs. 3 and 6. The similarity in both words and duration of the trajectories is striking. This similarity was repeated for each speaker's data across the dataset.

2. Duration of F_3 trajectory

The consistency of trajectory durations was tested statistically as follows. The F_3 duration values were entered into analyses of variance using the factors speaker (HD, RD, WJ, JM, SS, MS, BS), stress (initial or final), and context (/b/, /d/, /g/, /v/, or /r/). The hypotheses being considered were (1) whether F_3 trajectory duration differs as a function of stress condition, and (2) whether F_3 trajectory duration differs across contexts (speaker-to-speaker differences were expected). Because of correlations naturally existing across data from particular subjects, particular items, and particular stress patterns, we elected to treat each of these factors as a correlated variable in a repeated measures analysis of variance. Separate “subject” and “item” repeated measures analyses of variance were performed using, respectively, subject variability, consonant context variability, and stress variability as the error term. The subject analysis used context and stress as “within,” or “repeated” measures while speaker was a “between” or “grouping” factor. The items analysis used context and stress as “between” variables while subject variability was a “within” measure. Because

for subject RD data from /br/ was not collected, the items analysis context factor included /vr/, /dr/, /gr/, and /r/. The dependent variable in all cases was duration of the *F3* trajectory. For both sets of analyses, individual cells were represented for each analysis by means across tokens for a particular subject×context×stress combination. (Standard errors of tokens within all combinations exhibited a range of 0–25 ms, with a mean standard error of 9.96 ms.)⁴

Overall, context was not significant in either analysis (subject: $df=4, 23$, $F=1.99$, $p>0.10$; item: $df=3, 3$, $F=0.161$, $p>0.10$), suggesting that *F3* trajectory duration was consistent regardless of whether /r/ was the single intervocalic consonant, or whether it followed /b/, /v/, /d/, or /g/. Stress was significant in both subject and item analyses (subject: $df=1, 6$, $F=28.8$, $p<0.01$; item: $df=1, 3$, $F=99.8$, $p<0.001$), indicating that measured *F3* trajectory duration was different according to stress pattern. The interaction of stress×context, tested in the subject analysis, was not significant (subject: $df=4, 23$, $F=1.5$, $p>0.10$), suggesting that the effect of stress pattern was consistent across categories of items. There was a significant effect of speaker (item: $df=6, 18$, $F=23.8$, $p<0.001$) but interactions between context and speaker (item: $df=18, 18$, $F=1.1$, $p>0.10$) or stress and speaker were not significant (item: $df=6, 18$, $F=1.8$, $p>0.10$), suggesting that although speaker identity affected the duration of measured *F3* trajectories, these effects were consistent across all other variables. The speaker effect reflected characteristically longer or shorter *F3* trajectories for different speakers, presumably relating to intrinsic differences between subjects in terms of tongue musculature, mouth size, speech motor habits, etc. Subject-to-subject differences for singleton /r/ (/warav/) words, for instance, ranged from speaker BS's 206 ms to speaker HD's 264 ms. The overall mean across all subjects and all contexts was 224 ms.

The stress effect appeared to be due to a tendency for the *F3* trajectory edge detection algorithm to find a longer *F3* trajectory (by approximately 30 ms) in words whose initial vowel was unstressed. Trajectory length was positively correlated, across speakers and tokens, with the degree to which an unstressed initial /a/ vowel was reduced (to /ə/). The longest trajectory measurements were seen for speakers whose natural *F3* in back and central vowels was relatively low. Because identification of the /r/ trajectory beginning was dependent on the degree of lowering from /a/, it is not clear how much of the stress effect is attributable to expansion of the /r/-related *F3* trajectory and how much to difficulty in automatic identification of a relatively nonsalient inflection point. The fact that the slope of *F3* lowering and trajectory shape was extremely consistent for all speakers across stress suggests the latter (see Fig. 7).

As noted above, the context factor did not reach significance in either subject or item analyses, suggesting that trajectories for /r/ in labial, alveolar, and velar contexts were similar to those in singleton /r/ words. There were no significant interactions between speaker and context or between context and stress, indicating that the effects of context and stress were similar across subjects. Separate subject and item analyses excluding singleton /r/ words also showed no sig-

nificant effect of context; that is, there were no significant differences in trajectory duration between consonant contexts /b/, /v/, /d/, or /g/. There was a (nonsignificant) trend in the data for measured singleton /r/ trajectories to be shorter than those for /Cr/ words by approximately 20 ms.⁴ Because trajectories for /warav/ words were visually similar to those with /Cr/ consonants (random groupings of trajectories appearing very like those pictured in Figs. 7 and 8), we attribute the slightly longer measured trajectory durations in /Cr/ words (versus singleton /r/ words) to the existence of the consonant closure interval and consequent measurement artifact due to interpolation (see Sec. II A above). (The number of tokens where *F3* in the closure interval was sufficiently noisy to require interpolation, and the duration of interpolated regions was approximately the same for different consonant contexts pooled across speakers, although each speaker's pattern was different.)

3. Effects beyond trajectory edges

As Figs. 7 and 8 indicate, there is some lowering of *F3* before trajectory beginning as identified by the automatic procedure. It is possible that such lowering is anticipatory in nature as predicted by the feature spreading model; i.e., the time at which lowering begins may expand and contract according to context. Anticipatory lowering of this type would be expected to differ according to the identity of the consonant before /r/; again, earlier lowering would be expected when the intervening consonant was labial, while less and/or later lowering would be expected when the intervening consonant was alveolar or velar. Alternatively, lowering before trajectory edge may be part of a stable articulatory complex of movement for /r/. To test this question, formant values for /Cr/ and control words with singleton consonants were compared. Formant values were measured at syllable peaks as determined by an automatic procedure that identified the leftmost energy maximum in a 640–2800 Hz band and averaged the frequency value at this point with the values of the preceding and following frames. This method identified reliable formant values in a region both close to trajectory edge and salient for vowel perceptual identity. Syllable peak time points are illustrated by arrows in Fig. 3.

If the lowering we see before trajectory edge is under way by initial syllable peak, we might expect that formant values for control words would be slightly higher than those for /Cr/ words. This will be true whether the lowering reflects part of a relatively consistent, stable /r/-related movement, or if it reflects spreading of the /r/-related movement into preceding segments. However, the feature spreading model predicts more lowering when the consonant context is labial than alveolar or velar. If the lowering does not take place during the syllable peak, but after it (i.e., if /r/-related lowering does not start until after the syllable peak, we expect formant values for control words and /Cr/ words to be the same. The coproduction model predicts that lowering may or may not occur during the initial syllable peak, depending on the placement of the stable /r/ trajectory relative to the rest of the word. The amount of any lowering found would also be dependent on placement of the trajectory, which might vary according to context. Thus, a finding of

lowering by itself is compatible with both coarticulation models. However, a finding of consistency in the amount of lowering across different contexts is most compatible with the coproduction model.

The amount of lowering prior to initial syllable peak was tested statistically as follows. Formant values were entered into subject and item repeated measures analyses as described above for trajectory duration testing, using the factors stress (word-initial syllables were stressed or unstressed according to word stress condition), rhotic (control words versus words containing /Cr/ clusters), consonant (/b/, /d/, /g/, or /v/) and speaker (subject). Because of missing /b/ context data from subject RD, two items analyses were performed, one using only /d/, /g/, and /v/ contexts and one that repeated /v/ data for /b/ context in /b/ cells. The pattern of results was the same; only the latter analysis is reported below. Both analyses used means across tokens for subject \times context \times stress \times rhotic combinations. (Standard errors for tokens within cells ranged from 1–141 Hz, with a mean standard error of 37 Hz.⁵) Overall, the effects of stress and rhotic were significant in both the subject and items analysis; stress (subject: $df=1, 5, F=10.9, p<0.05$; item: $df=1, 10, F=111.8, p<0.001$), rhotic (subject: $df=1, 5, F=10.5, p<0.05$; item: $df=1, 10, F=31.9, p<0.001$). These results were due to overall higher $F3$ at the measurement point for stressed /a/ vowels, probably due to reduction during unstressed vowels, and overall lower $F3$ at the measurement point for vowels in /Cr/ words, probably due to proximity to the $F3$ trajectory. The mean difference between $F3$ at syllable peak for /Cr/ versus control words was 53.4 Hz. The mean difference between $F3$ at syllable peak for initial stress versus final stress words was 103.9 Hz. Speaker was not a separate variable in the subject analysis (data entries being treated as correlated) but was significant in the items analysis (item: $df=6, 60, F=16.6, p<0.001$). This result was expected given differences in vocal tract geometry for different speakers. The main effect of consonant was significant as well (subject: $df=3, 15, F=4.3, p<0.05$; items: $df=3, 10, F=6.3, p<0.05$), indicating that proximity to different consonants affects $F3$ during the preceding vowel. Among interactions, only those between speaker and variables stress and rhotic were significant; stress \times speaker ($df=6, 60, F=5.1, p<0.01$); rhotic \times speaker ($df=6, 60, F=2.6, p<0.05$). (Interactions with speaker were assessed in the items analysis only, as speaker was not a main variable in the subject analysis.) The interaction consonant \times speaker was not significant ($df=18, 60, F=1.1, p>0.05$). The interaction of stress with speaker was expected, given that different speakers had different patterns of vowel reduction. The interaction of rhotic with speaker indicates different amounts of lowering (in Hz) for /r/ during the preceding vowel. This could be due to the fact that different speakers have different formant values for $F3$ during /a/, and thus lower less or more for /r/ as a normal part of the /r/ trajectory. In addition, we know that speakers differ in the overall duration of their trajectories; if these trajectories have stable durations across context they are likely to begin at different points in the vowel (a plausible version of the coproduction model). Alternatively, it could reflect changes in the consonant \times rhotic interaction across

speakers. This would be a plausible variant of the feature spreading model. However, such an effect should be echoed in a significant consonant \times rhotic interaction and/or a significant consonant \times rhotic \times stress interaction. In this study, interaction effects between nonspeaker variables were assessed in the subject analysis; none were significant; stress \times rhotic (subject: $df=1, 5, F=0.23, p>0.05$), consonant \times stress (subject: $df=3, 15, F=0.18, p>0.05$) and consonant \times rhotic (subject: $df=3, 15, F=0.44, p>0.05$), stress \times consonant \times rhotic ($df=3, 15, F=0.5, p>0.05$) (Similar results were obtained in a 3-factor ANOVA treating subject and item identity as uncorrelated variables, with the exception that in this analysis no interactions were significant.) The lack of interaction effects with consonant suggests that early lowering, like the duration of the bulk of the trajectory itself, is not affected by consonant context.

III. EXPERIMENT 2: ARTICULATORY MOVEMENT FOR /r/

A. Methodology

To get some idea of correspondence between $F3$ trajectory and articulatory movement of the tongue, and to confirm the validity of the $F3$ trajectory duration measure, articulatory plus acoustic data were obtained from one speaker, RD, who produced a subset of the experimental corpus listed above. These articulatory data from RD were used to confirm: (a) that the acoustic time course of $F3$ represents the articulatory time course of (primary constriction) tongue movement for /r/, (b) that use of the “lower” rather than “upper” path for this subject, and, by analogy, other subjects, is the appropriate acoustic index of /r/, and (c) that articulatory trajectories show shape and duration consistency similar to that found for $F3$ trajectories. Because of the difficulty of visually or algorithmically separating tongue movement for /r/ from that for /g/ and /d/, only the comparison between singleton /r/ and /vr/ contexts is shown here.

The corpus included all nonsense words with the exception of /wabav/ and /wabrav/ as well as real words with parallel structure. Movement of two electromagnetic transducers placed on the tongue tip and tongue dorsum was recorded via an Electro-Magnetic Midsagittal Articulometer (EMMA) apparatus (Perkell *et al.*, 1992). The acoustic signal was recorded with a directional microphone. The subject was seated in a quiet room and produced the experimental stimuli in randomized order by visual reference to a list suspended at eye level. The subject’s head was restrained from movement by a specially designed headpiece. Transducers were attached to the subject’s tongue at approximately 1 and 5 cm from the apex of the tongue tip along the tongue midline. Again, the speaker was instructed to maintain a consistent speaking rate. Movement in the anterior-posterior (X) and superior-inferior (Y) dimension was recorded separately for the tongue tip (TT) and the tongue dorsum (TD) transducer. Movement signals were digitized simultaneously with the audio signals at a rate of 312.5 Hz. Acoustic formant track data and movement signal frame rates were matched by recomputing the acoustic formant tracks with a 51.2-ms window and a 3.2-ms frame rate. These were aligned by taking

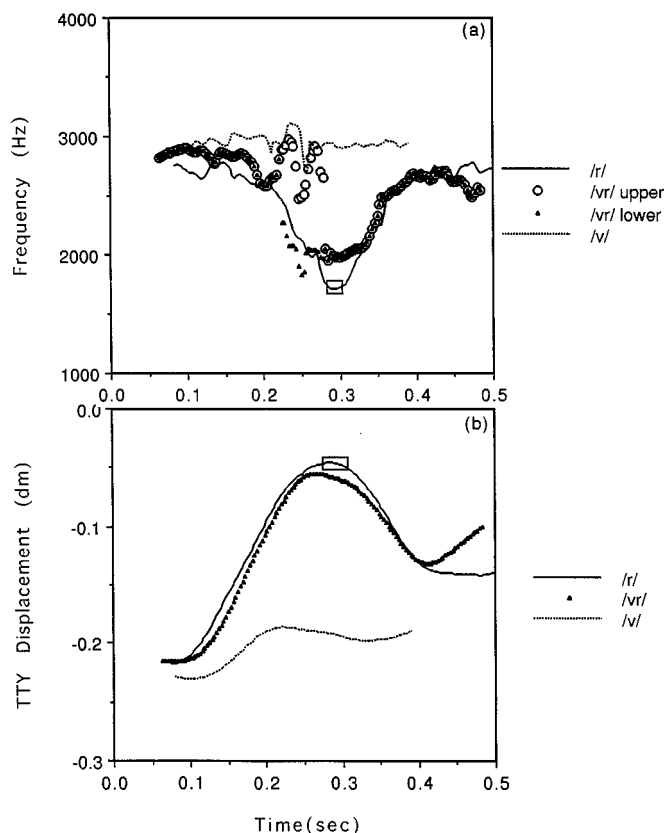


FIG. 9. (a) F_3 tracks shown in part (c) of Fig. 5 for speaker RD. (b) Corresponding articulatory data showing the time course of the upward (superior-inferior) movement of the tongue tip during the /ara/ of /'warav/, the /avra/ of /'wavrav/, and the /ava/ of /'wavrav/.

into account a shift of half the window length (i.e., eight sample points).

B. Correspondence of acoustic and articulatory time course for singleton /r/ words

Subject RD generally raised his tongue tip to make the primary constriction for /r/ [except for the /gr/ context, where tongue dorsum was raised (Espy-Wilson and Boyce, 1994)]. The correspondence between articulatory movement and F_3 trajectory for /r/ proper was explored in detail as follows: For 11 tokens of intervocalic /r/ (six tokens of /warav/, two tokens of "pariah," two tokens of "barometer" and one token of "bar"), maximum tongue tip transducer (TTY) and minimum F_3 values were determined by automatic procedure. These were an average of 21.5 ms apart in time, with the F_3 minimum occurring after the TTY maximum in all cases (range: 6–38 ms). When measurement error was taken into account (i.e., ranges of TTY or F_3 points to the left and right of extrema within a measurement error of 0.3 mm for TTY and 10 Hz for F_3),⁶ the TTY maximum overlapped with the leftmost F_3 minimum in all cases. This is illustrated in Fig. 9, which shows the acoustic F_3 trajectory for tokens of /'warav/, /'wavrav/, and /'wavrav/ together with time-aligned superior-inferior (Y) movement of the tongue tip. Ranges for F_3 minimum and TTY maximum are indicated by rectangular boxes. As in previous figures, F_3 and TTY values start at the beginning

of the stressed /a/ and end with the end of the unstressed /a/. It is clear from Fig. 9 that the acoustic F_3 trajectory and the TTY trajectory parallel each other. No clear demarcation exists in the articulatory TTY data to correspond with the inflection points on the F_3 trajectory, i.e., where F_3 became unambiguously associated with /r/, but the tongue tip and F_3 tracks show broad peaks and valleys occurring at congruent points in time. These data confirm that the F_3 trajectory is a reasonable index of articulatory movement for /r/.

Articulatory data from /Cr/ tokens also confirmed that movement related to /r/ was reflected in "lower" rather than "upper" path F_3 trajectories. This is shown in Fig. 9, which combines the acoustic trajectories shown in Fig. 5 with their associated TTY movement. As noted above, the TTY maxima and F_3 minima for the "lower path" occur at synchronous points in time. In contrast, the same comparison for the "upper path" trajectory in /'wavrav/ reveals no visible congruence of timing. Rather, the "upper path" resembles the TTY trajectory for /'wavrav/ in that both show little change from F_3 and tongue movement position for /a/. These data confirm that the acoustic "upper path" data reflect articulatory shaping of the vocal tract specific to /v/.

Altogether, congruent relationships between the "lower" F_3 trajectory track and maximum TTY movement (or maximum TDY, for /gr/ words) were characteristic of all real and nonsense words containing /r/ produced by subject RD. We interpret this trajectory congruence (F_3 lowering and TTY raising for /dr/, F_3 lowering and TDY raising for /gr/) as showing that, everything else being equal, "lower path" F_3 lowering in the acoustic domain is a reasonable index of the time course of articulatory movement specific to /r/.

IV. CONCLUSION

The data in this study demonstrate that F_3 trajectories for /r/, for any one subject, show relatively consistent duration and shape across a number of variables that might be expected to affect the way /r/ is articulated. Notably, the qualitative similarity in trajectory shape suggests that duration for all components of F_3 trajectories—early onset (lowering), extremum, and offset (raising)—remains consistent across phonetic context. Similarly, the measured duration of the full F_3 trajectory is consistent across phonetic contexts. Thus it appears that whether the segment preceding /r/ is alveolar, velar, labial, or vocalic does not affect the essential shape or duration of the F_3 trajectory. In a global sense, this result is more consistent with the coproduction model of coarticulation, which predicts that articulatory trajectories for a particular segment will tend to show stable profiles across segmental contexts, than with the traditional "feature spreading" model, which predicts that F_3 trajectories will change shape and duration (i.e., lengthen or shorten) according to the articulatory and acoustic requirements of the adjoining segments. Given the results reported here, it is unclear whether stress change accompanied by vowel reduction operates to change F_3 trajectory duration (primarily by flattening the lowering curve from the initial vowel). Note that an effect of this nature due to stress is consistent with both coarticulation models. However, the minimal effect of vowel

reduction and/or stress on trajectory shape seen in the present data suggests that any effect is minor at best.

As noted above, /r/ is an articulatorily complex segment, involving variant forms of tongue tip and tongue body movement as well as varying combinations of these with labial and pharyngeal narrowing. Subjects may have a number of strategies available to deal with articulatory difficulty in combining /r/ with alveolar or velar contexts. For instance, a speaker may alternate between use of /r/ variants so as to use the tongue dorsum for /r/ in alveolar contexts, and the tongue tip for /r/ in velar contexts (Espy-Wilson and Boyce, 1994). From our articulatory data, it appears that subject RD used compatible forms of retroflex /r/ in /dr/ contexts and bunched /r/ in /gr/ contexts. The acoustic evidence presented for the subjects examined here does not allow us to identify the precise articulatory mechanisms involved in producing /r/ for the remaining subjects, words, and contexts examined here. However, even if we assume that the subjects in this study used varied articulatory strategies, and that these strategies varied both idiosyncratically and by context, we might expect such variation in articulatory strategies to result in some degree of variability of acoustic patterns for /r/ coarticulation between speakers and across contexts. Certainly, we know of no external articulatory factors that would prevent /r/ variation in duration and shape over context. In view of this articulatory complexity, the consistency in duration and shape exhibited in this study by the acoustic *F3* trajectory for /r/ across contexts is notable. This is particularly the case when we consider the similarity between singleton /r/ and labial contexts, and between labial and lingual consonant environments. Altogether, these results suggest that articulatory movement for /r/ may be organized specifically to achieve a consistent acoustic pattern. In other words, the maintenance of acoustic (and articulatory) movement profiles over segmental context may be an organizing principle of the speech system.

This study started with an observation of consistency in *F3* trajectories, accompanied by speculation that much of what appears to be variability in the acoustic record is more apparent than real. Consequently, we suggest that much of what has been described as phonological and coarticulatory interaction between /r/ and surrounding segments can be attributed to trajectory overlap and “sliding” rather than “spreading” of /r/-related characteristics and attendant change in the articulatory plan for /r/. With regard to possible articulatory mechanisms of trajectory maintenance, we can only speculate. One possibility is that subjects “swap” between bunched and retroflex articulations of /r/ according to the requirements of context. Alternatively, the fact that overlapping constrictions for obstruents such as /g/ and /d/ interfered so little with *F3* trajectories for /r/ suggests that constriction location is less important than some other manipulation of the vocal tract affecting resonance. That is, perhaps the primary producer of /r/-related movement, and the source of its uniquely low *F3*, is not the tongue tip or tongue dorsum place of articulation *per se*, but simultaneous narrowing at some more peripheral portion of the vocal tract, or the formation of an extra resonating cavity, or some complex interaction between these factors. It will be necessary, in the

future, to expand these findings to additional data from real words, with different articulators, and with different numbers of syllables and different vocalic contexts.

ACKNOWLEDGMENTS

The articulatory data discussed in this paper were collected with the generous help of Dr. Joseph Perkell, Dr. Melanie Matthies, Dr. Mario Svirsky, and Jane Wozniak. We are indebted to them and to many others for comments and suggestions, including Dr. Kenneth Stevens, Nabil Bitar, Demetrios Paneras, Dr. Caroline Huang, Dr. Sharon Manuel, Dr. Rena Krakow, Dr. Patricia Keating, Dr. Daniel Recasens, Dr. Jan Edwards, one anonymous reviewer, and associate editor Dr. Anders Lofqvist. In addition, we thank Dr. Raymond Kent for sharing his recollection of past x-ray studies of /r/. This work was supported in part by: NSF Grants No. 9296032 and No. IRI-9310518, a Clare Boothe Luce fellowship to the second author, and a Boston University College of Engineering Dean's postdoctoral fellowship, and NIH Grant No. 1R030C-02576 to the first author.

¹It is notable that except for the vowel /i/, where *F3* is around 3000 Hz, and for rounded vowels, where *F3* may be as low as 2200 Hz, *F3* for vowels tends to remain within a 2300–2500 Hz band (Ladefoged, 1982; Zue, 1985). Individual speakers may vary proportionately; subjects BS and RD in the present study, for instance, had *F3*'s of 2700–3000 Hz during /a/, and trajectory “bends” some 300–500 Hz lower. Apart from /r/, obstruent-type constriction at specific points on the palate, in the pharynx, constriction at the lips, and/or extension of the vocal tract lengthwise such as occurs during lip protrusion, will lower *F3*, but probably by no more than 200–300 Hz (Kewley-Port, 1982; Espy-Wilson and Boyce, unpublished modeling study).

²In situations where there was no clear acoustic landmark separating a semi-vowel from an adjacent vowel, the heuristic rule applied for segmentation was to assign 2/3 of the vowel and semivowel region to the vowel and the remainder to the semivowel.

³HSS and MS had participated in an earlier pilot study of four subjects whose results resemble those produced here. At that time, neither were aware of the purpose of the study. BS was a researcher associated with the study.

⁴A three-factor analysis of variance using tokens, in which speaker- and item-specific characteristics were treated as uncorrelated variables, yielded a similar pattern of results. The one exception was a significant difference between singleton /r/ and /Cr/ words. This effect was traced to an interaction between context and subject, whereby speakers WJ and HD showed longer measured trajectories in /Cr/ words (see Sec. II A 2 for a possible explanation). This effect was not significant in the repeated measures analysis, which takes account of intraspeaker correlations.

⁵As with the duration measures, a three-factor analysis of variance using tokens, in which speaker- and item-specific characteristics were treated as uncorrelated variables, yielded a similar pattern of results.

⁶Frequency resolution for the ESPS/WAVES formant tracker in the *F3* frequency band was empirically determined using synthesized /a/ and /r/. When an all-pole model was assumed, frequency was matched to within 2 to 3 Hz. When one antiresonance was included in the four-formant model, frequency was matched to within 55 Hz. Since whether the acoustics of /r/ includes an antiresonance is not known, we used 10 Hz as a conservative compromise.

Alwan, A., Narayanan, S., and Haker, K. (1997). “Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II: The rhotics,” *J. Acoust. Soc. Am.* **101**, 1078–1089.

Bell-Berti, F., and Krakow, R. (1991). “Anticipatory velar lowering: A coproduction account,” *J. Acoust. Soc. Am.* **90**, 112–123.

Bell-Berti, F., Krakow, R. A., Gelfer, C. E., and Boyce, S. E. (1995). “Anticipatory and carryover effects: Implication for models of speech production,” *Producing Speech: A Festschrift in honor of Katherine Safford Harris*, 77–97, edited by F. Bell-Berti and L. Raphael (American Institute of Physics, Woodbury, NY).

- Bernthal, J., and Bankson, N. (1993). *Articulation and Phonological Disorders* (Prentice-Hall, Englewood Cliffs, NJ), pp. 5–60.
- Boyce, S. E., Krakow, R. A., Bell-Berti, F., and Gelfer, C. (1990). “Converging sources of evidence for dissecting articulation into core gestures,” *J. Phon.* **18**, 173–188.
- Bronstein, A. (1967). *Your Speech and Voice* (Random House, New York).
- Browman, C. P., and Goldstein, L. M. (1986). “Towards an articulatory phonology,” *Phonology Yearbook* **3**, 215–252.
- Browman, C. P., and Goldstein, L. M. (1990). “Gestural specification using dynamically-defined articulatory structures,” *J. Phon.* **18**, 299–320.
- Daniloff, R., and Moll, K. (1968). “Coarticulation of lip-rounding,” *J. Speech Hear. Res.* **11**, 707–721.
- Delattre, P. (1967). “Acoustic or articulatory invariance?,” *Glossa*, **1**, 3–25.
- Delattre, P., and Freeman, D. (1968). “A dialect study of American r’s by x-ray motion picture,” *Language* **44**, 29–68.
- Espy-Wilson, C. Y. (1987). “An acoustic-phonetic approach to speech recognition: Application to the semivowels,” Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Espy-Wilson, C. Y. (1992). “Acoustic measures for linguistic features distinguishing the semivowels in American English,” *J. Acoust. Soc. Am.* **92**, 736–757.
- Espy-Wilson, C. Y. (1994). “A feature-based semivowel recognition system,” *J. Acoust. Soc. Am.* **96**, 65–72.
- Espy-Wilson, C., and Boyce, S. (1994). “Acoustic differences between “bunched” and “retroflex” variants of American English /r/,” *J. Acoust. Soc. Am.* **95**, 2823(A).
- Espy-Wilson, C., and Boyce, S. (unpublished).
- Fowler, C. (1993). “Phonological and Articulatory Characteristics of Spoken Language,” in *Linguistic Disorders and Pathologies: An International Handbook*, edited by G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, and C.-W. Wallesch (Walter de Gruyter, New York), pp. 34–46.
- Gelfer, C., Bell-Berti, F., and Harris, K. (1989). “Determining the extent of coarticulation: Effects of experimental design,” *J. Acoust. Soc. Am.* **86**, 2443–2445.
- Giergerich, H. (1992). *English Phonology: An Introduction* (Cambridge U.P., Cambridge, England).
- Gracco, V., and Lofqvist, A. (1994). “Speech motor coordination and control: Evidence from lip, jaw and laryngeal movements,” *J. Neurosci.* **14**(11), 6585–6597.
- Hagiwara, R. (1995). “Acoustic realizations of American English /r/ as produced by women and men,” *UCLA Phonetics Laboratory Working Papers* **90**, 55–61.
- Hammarberg, R. (1976). “The metaphysics of coarticulation,” *J. Phon.* **4**, 353–363.
- Harris, K., and Bell-Berti, F. (1984). “On consonants and syllable boundaries,” in *Language and Cognition*, edited by L. J. Raphael and M. R. Valdovinos (Plenum, New York).
- Keating, P. (1988). “Underspecification in phonetics,” *Phonology* **5**, 275–292.
- Kent, R., and Minifie, F. (1977). “Coarticulation in recent speech production models,” *J. Phon.* **5**, 115–133.
- Kewley-Port, D. (1982). “Measurement of formant transitions in naturally produced stop consonant-vowel syllables,” *J. Acoust. Soc. Am.* **73**, 379–389.
- Ladefoged, P. (1982). *A Course in Phonetics* (Harcourt Brace Jovanovich, San Diego).
- Lehiste, I. (1962). “Acoustical characteristics of selected English consonants,” University of Michigan Communication Sciences Laboratory Report No. 9.
- Lehiste, I., and Peterson, G. E. (1961). “Transitions, glides, and diphthongs,” *J. Acoust. Soc. Am.* **33**, 268–277.
- Lindau, M. (1985). “The story of /r/,” in *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, edited by V. Fromkin (Academic, Orlando).
- Munhall, K., and Lofqvist, A. (1992). “Gestural aggregation in speech: Laryngeal gestures,” *J. Phon.* **20**, 111–126.
- Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition* (Cambridge U.P., Cambridge, England).
- Olive, J., Greenwood, A., and Coleman, J. (1993). *Acoustics of American English Speech* (Springer-Verlag, New York).
- Perkell, J. S., and Matthies, M. L. (1992). “Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability,” *J. Acoust. Soc. Am.* **91**, 2911–2925.
- Perkell, J. S., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (1992). “Electro-Magnetic Midsagittal Articulometer (EMMA) systems for transducing speech articulatory movements,” *J. Acoust. Soc. Am.* **92**, 3078–3096.
- Peterson, G. E., and Barney, H. L. (1952). “Control methods used in a study of the vowels,” *J. Acoust. Soc. Am.* **24**, 175–184.
- Recasens, D. (1985). “Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences,” *Language Speech* **28**, 97–114.
- Seneff, S., and Zue, V. (1988). “Transcription and alignment of the TIMIT database,” documentation distributed with the TIMIT database by NBS.
- Shoup, J., and Pfeifer, L. (1976). “Acoustic characteristics of speech sounds,” in *Contemporary Issues in Experimental Phonetics*, edited by N. Lass (Academic, New York).
- Westbury, J. M., Hashi, M., and Lindstrom, M. J. (1995). “Differences among speakers in articulation of American English /r/: An x-ray microbeam study,” in *Proceedings of the XIIIth International Conference on Phonetic Sciences*, 1995, Vol. 4, 50–57.
- Zue, V. (1985). *Speech Spectrogram Reading*, Summer Course, MIT.

The characteristics of voicing in syllable-initial fricatives in American English

Karen Pirello,^{a)} Sheila E. Blumstein, and Kathleen Kurowski

Brown University, Department of Cognitive and Linguistic Sciences, Box 1978, Providence, Rhode Island 02912

(Received 27 February 1996; accepted for publication 8 January 1997)

This study investigated the acoustic characteristics of voicing in the production of fricative consonants. The fricatives [f v s z] were used in combination with the vowels [i e a o u] to create CV syllables, which were produced by four subjects both in a context condition (following voiced and voiceless velar stops) and in isolation. Analyses were conducted of the time course of glottal excitation during the fricative noise interval in the voiced and voiceless fricative stimuli. Results showed that the patterns of voicing in the fricative noise interval were influenced by the voicing characteristics of preceding stop consonants. Nonetheless, these carryover coarticulatory effects were short-lived, influencing only the first 10's of ms of the following segment. Despite the influence of phonetic context on the patterns of voicing, an acoustic measure relating to the presence or absence of glottal excitation at the acoustic boundaries of the fricative noise reliably classified a majority (93%) of the fricative consonants in terms of the phonetic category of voicing. Thus, while phonetic context affected the patterns of glottal excitation in the fricative noise interval, it did not affect the criterial attribute associated with the phonetic category of voicing. © 1997 Acoustical Society of America. [S0001-4966(97)04405-6]

PACS numbers: 43.70.Fq, 43.70.Hs, 43.72.Ar [AL]

INTRODUCTION

A number of acoustic parameters have been identified which contribute to the voicing distinction in fricative consonants. They include the presence of periodicity during the fricative noise interval, the intensity and duration of the fricative noise, fundamental frequency patterns in the following vowel, and the duration of formant transitions into the vowel (Slis and Cohen, 1969; cf. Lehiste, 1970 for discussion). Nonetheless, the acoustic parameter that seems to be most salient from a perceptual point of view is the presence of vocal cord activity during the fricative noise interval (Stevens *et al.*, 1992). Although Denes (1955) reported that syllable final fricatives were perceived as voiced when their durations were shortened and as voiceless when their durations were lengthened, Jongman (1989) showed that shortening the duration of the frication noise in initial voiceless and voiced fricatives did not change the perception of voicing, even when the duration of the frication noise was reduced to 30 ms.

Investigations focusing on the acoustic evidence for the presence of vocal cord activity show that in voiced fricatives glottal excitation does not necessarily occur throughout the duration of the frication noise. Rather, considerable devoicing occurs, and it seems to be affected by contextual factors such as syllable position, place of articulation of the fricative itself, and voicing characteristics of adjacent consonants (Haggard, 1978; Stevens *et al.*, 1992). The major conclusion drawn from these studies is that the feature voicing in fricatives is manifested in a continuous way and as such cannot

be characterized in terms of a binary distinction relating to the presence or absence of glottal excitation. As Haggard states, "... in fluent speech, voiced fricative production is a dynamic trajectory through a voiced phase and a fricative phase, with little effort given to the synchronization of these phases" (1978, p. 96).

Nonetheless, Stevens *et al.* (1992) showed that, despite the variability inherent in the glottal excitation in the frication noise for voiced fricatives, the synchronization of voicing with respect to frication does seem to be a critical attribute. In particular, 20–30 ms of glottal excitation was present during the fricative interval at either the boundary of the consonant implosion or at the release in these consonants. This acoustic measure accurately classified approximately 83% of voiced and voiceless fricatives in the singleton intervocalic position and in intervocalic clusters matched for voicing. The measure was less successful in categorizing utterance final fricatives and fricative clusters with mixed voicing.

Based on the results of the studies just reviewed, it is clear that voicing in fricative consonants is variable and is affected by a number of contextual factors. The question remains whether this variability is in itself systematic and whether, despite this variation, a common property can be derived to characterize the phonetic category corresponding to voicing in fricative consonants. The current study was designed to explore these issues and to extend the work of Haggard (1978) and Stevens *et al.* (1992).

Two questions were investigated. The first question explored whether systematic patterns of voicing could be identified as a function of phonetic context for both voiced and voiceless fricatives. Haggard's results suggest that the voicing of stop consonants influenced the patterns of voicing in

^{a)}Now at DP Solutions, Inc., 4249 Piedmont Parkway, Suite 105, Greensboro, NC 27410.

neighboring fricative consonants. However, in his study, voiced and voiceless stops did not occur in the same phonetic position relative to the fricative consonants, making it difficult to quantify and compare the influence of consonant voicing on the production of voicing in fricatives. Moreover, only voiced fricatives were analyzed. It would be worthwhile to determine whether context effects emerge not only for voiced fricatives but for voiceless fricatives as well. That is, would a preceding voiced stop affect the pattern of glottal excitation in the production of voiceless fricatives? If such results are found, would voiceless fricatives no longer be characterized phonetically as voiceless? If this were the case, these findings would add to the complexity of the voicing dimension for fricative consonants. The results of Stevens *et al.* (1992) suggest that indeed voicing in voiceless fricatives is influenced by the presence of voicing in neighboring fricatives. They showed an asymmetric effect with less influence of voiced fricatives on voiceless fricatives than voiceless fricatives on voiced fricatives.

The second question we intend to investigate is whether an acoustic property for voicing in fricatives can be identified that remains stable despite the types of variability described above. This question derives from certain assumptions about the nature of the phonetic features (i.e., phonetic categories) of speech. In particular, it has been proposed that for any given phonetic feature, there is an acoustic property or set of acoustic properties that define that phonetic feature (Stevens and Blumstein, 1981). It is further assumed that an acoustic property associated with a phonetic feature encompasses a range or “family” of values that allow for the acoustic manifestation of a feature to vary within certain definable limits. For example, there is a family of spectra that corresponds to the diffuse-rising spectral shape associated with the alveolar place of articulation in stop consonants (Blumstein and Stevens, 1979), and there is a range of amplitude values in the frication noise that corresponds to the acoustic property associated with the feature strident (Utman and Blumstein, 1994). With respect to voicing in fricative consonants, there is a range of values encompassing both the amplitude and duration of glottal excitation that corresponds to the acoustic property proposed for voicing in fricative consonants (cf. Stevens *et al.*, 1992, and elaborated in Sec. I). Where a particular voiced or voiceless fricative may fall within this range and the patterns of glottal excitation that may occur is presumed to be influenced by a number of factors including articulatory imprecision in the implementation of speech as well as coarticulatory effects. Despite this variability, the criterial attribute such as the one proposed by Stevens *et al.* (1992) of 20–30 ms of voicing during the fricative noise interval should remain present across these sources of variability. In this sense, the acoustic property associated with voicing in fricative consonants remains stable.

Stevens *et al.*’s measure seems to provide a promising first approximation for determining whether a stable acoustic property can be identified for voicing in fricative consonants. The measure successfully characterized 83% of voiced and voiceless fricatives in a number of environments. The categorization of voicing was not as successful in some other

environments. As Stevens *et al.* (1992) discuss, the poor performance of their measure in categorizing voicing in fricative clusters with mixed voicing (where correct categorization fell to 53% for voiced fricatives) could have reflected the difficulties inherent in identifying the boundaries between the two members of fricative–fricative clusters, the presence of true voicing assimilation in the cluster environments (hence a high level phonological planning effect, not a phonetic implementation effect), and/or phonetic errors made by the speakers who were required to produce articulatorily complex utterances, e.g., *his vaf sips it*. Thus, it is possible that a measure related to the presence/absence of glottal excitation at the acoustic boundaries or areas of acoustic discontinuity provides a reliable and stable measure of voicing in fricative consonants. Such results would support the view that there is an acoustic property corresponding to the phonetic category of voicing, and this property remains stable across numerous sources of variability inherent in the speech stream.

To explore whether such a measure could be identified, we focused on the production of voiced and voiceless fricatives in syllable initial position following voiced and voiceless stop consonants. The acoustic boundaries between stops and fricatives are easy to identify and allow for a straightforward analysis in charting the relation between the frication noise and glottal excitation. Based on the results of previous studies, we would expect to find residual glottal excitation in voiceless fricatives following a voiced stop, and conversely, some devoicing of voiced fricatives following voiceless stops. Nonetheless, despite the different patterns of voicing that may emerge, it is our expectation that an acoustic measure relating to the presence of glottal excitation at the acoustic boundaries of the fricative noise (either at the onset of the frication noise or in the interval of the frication noise immediately preceding the onset of the vowel) will reliably categorize voicing in fricative consonants. To quantify the changes that are made as a function of context, three conditions are included: an isolation condition in which fricative consonants are produced in syllable initial position followed by a vowel, and two context conditions, in which the same syllables are produced, but embedded in an utterance containing a preceding voiced or voiceless stop consonant.

I. METHODS

A. Subjects

Two adult male and two adult female speakers of American English served as subjects. All four subjects were trained in phonetics. Individuals trained in phonetics were used as subjects to facilitate (and to avoid training in) the reading of the fricative-vowel stimuli using phonetic transcriptions. There was no expectation that individuals trained in phonetics would produce the test utterances any differently from those not trained in phonetics.

B. Stimuli

There were two sets of stimuli: one set for the *isolation* condition, and one set for the *context* condition. The isolation

condition served as a baseline so that potential differences in patterns of voicing could be compared as a function of phonetic context.

In the isolation condition, the stimuli were CV syllables consisting of the four fricative consonants [f v s z] followed by one of five vowels [i e a o u]. Syllables were presented in random order, listed in phonetic form, for reading by the subjects. Each syllable occurred eight times. The subjects were instructed to read the list of fricative–vowel syllables in citation form (avoiding “list” intonation), and to speak clearly but naturally. Each subject produced a total of 160 fricative–vowel syllables.

In the context condition, fricative–vowel syllables were read in two different sentence contexts containing different phonetic environments: (1) following a voiceless velar stop consonant (“Please speak—again”), and (2) following a voiced velar stop consonant (“Say a big—again”). The sentences were printed in random order, each sentence appearing eight times. Subjects were instructed to read the sentences clearly and carefully, but as naturally as possible. Each subject produced a total of 320 sentences.¹

C. Procedure

Stimuli were recorded onto magnetic tape in a sound-treated room using a Nagra 4.2 recorder and a Shure condenser microphone. The microphone was placed about 20 cm from the speaker’s mouth. The stimuli were then digitized onto a VAX Station II at a sampling rate of 20 kHz with a 9.0-kHz low-pass filter (Butterworth 24 dB/octave), and 12-bit quantization.

D. Analysis procedures

The fricative–vowel stimuli were listened to by two of the experimenters (KP and KK). All of the fricatives were perceived to be members of their target phonetic categories.

To begin the acoustic analyses, cursors were placed at the onset and offset of the fricative noise for each fricative–vowel syllable using a software waveform editor (see Fig. 1 and also the spectrogram in Fig. 2). The beginning of the frication noise was identified on the waveform display by the presence of high-frequency noise, and was confirmed by DFT and LPC analyses (using a 15-ms full-Hamming window) as an increase of at least 10 dB in the high-frequency range of about 6–10 kHz, relative to the background noise level. In the case of fricatives produced in the context condition, the onset of noise was identified in the waveform display following the closure for [k] and [g]. In a few instances, there was frication noise coincident with the preceding velar release, and thus the onset of the fricative was placed at the point where the frication noise was identified. If an epenthetic vowel preceded a voiced fricative, it was not included as part of the frication noise measure unless frication was observed by our criterion. An epenthetic vowel can be distinguished from prevoicing in the fricative by the presence of formant structure above 800 Hz in the former and its absence in the latter.

The offset of the noise was determined by the absence of high-frequency noise in the waveform display, by the emer-

gence of a harmonic spectrum for the following vowel, especially at mid and high frequencies, and also by a greater than 10-dB increase in spectral energy (measured by DFT) at or above 2 kHz. For voiceless fricatives, the end of frication and the beginning of periodicity of the vowel were easily identified by visual inspection. The offsets of voiced fricatives were somewhat more difficult to establish. In these cases, a combination of DFT and LPC analyses was used to identify when the vowel’s formant structure had emerged. A concurrent drop of at least 10 dB in the frequency range of about 6–10 kHz was used as a secondary, supporting measure indicating the termination of the frication noise.

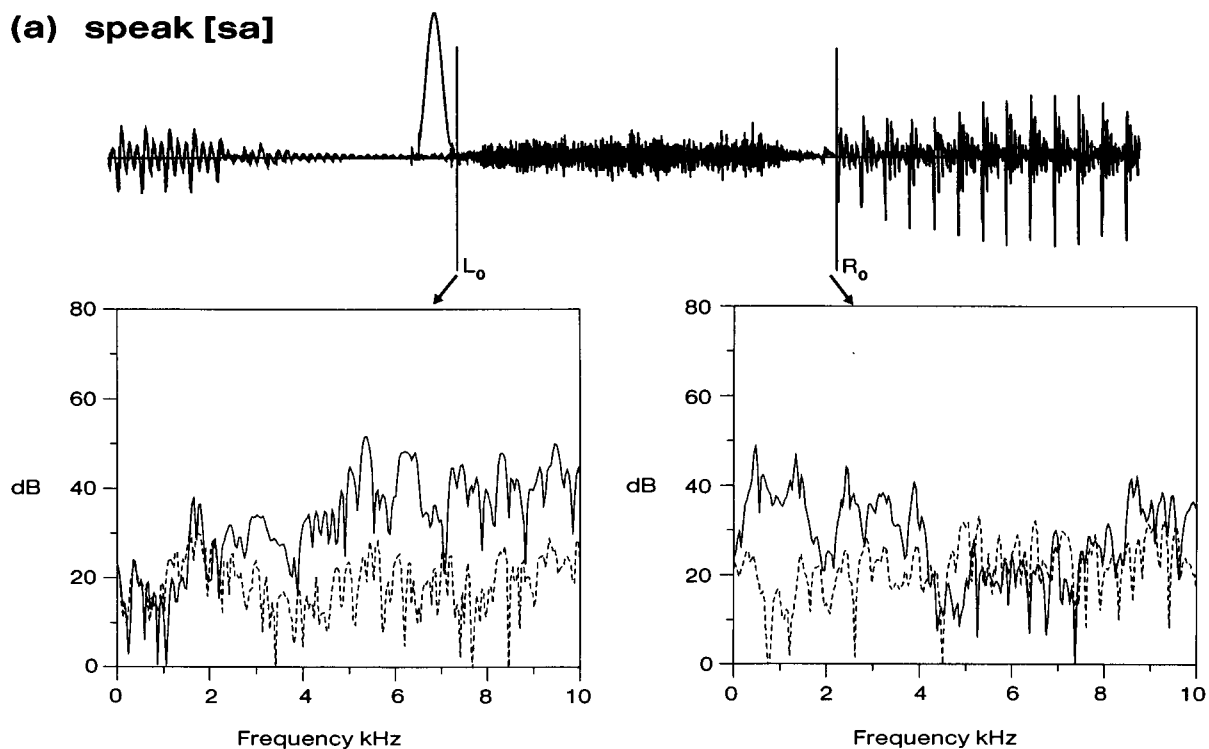
The cursor placements were corroborated by a second experimenter on a selected subset of utterances from each of the four speakers. There was approximately 90% agreement on the cursor placements.

After the fricative noise was demarcated, two measures were taken. The first was the duration of the frication noise determined by the time interval between the cursor settings. The second was a measure of voicing. The amplitude of the first harmonic, used as a measure of the amplitude of the glottal pulse (cf. Stevens *et al.*, 1992), was charted using a frame by frame analysis throughout the duration of the frication noise (see Fig. 2). Specifically, a 30-ms full-Hamming window was advanced in 10-ms steps throughout the frication noise. The amplitude of the first harmonic (*H1*) under each window was taken as a measure of voicing at that point in the fricative (Stevens *et al.*, 1992). These measurements were made until the window had passed the cursor marking the end of the frication noise, advancing into the onset of the following vowel to the point when the window included only the vowel, and excluded any frication noise. This last window position was used to ascertain the amplitude of *H1* for the vowel.²

Voicing in the frication noise interval was defined with reference to the amplitude of the following vowel (see bottom right panel of Fig. 2). The amplitude of *H1* in the frication noise relative to that of the following vowel served as an operational definition of voicing. In particular, for each fricative, the amplitude of *H1* for each 30-ms window was compared with the amplitude of the following vowel, i.e., the amplitude of *H1* in the 30-ms interval immediately following the offset of the frication noise. An amplitude difference greater than 10 dB between the amplitude of the vowel and fricative noise was classified as voiceless; a difference of less than or equal to 10 dB was classified as voiced. Following Stevens *et al.* (1992), a fricative was classified as “voiced” if at least 30 ms of contiguous voicing, i.e., three consecutive windows, was present at either the onset or offset of the frication noise; fricatives that failed to meet this criterion were categorized as “voiceless.”

We used a criterion of 0–10 dB sustained over 30 ms to provide a first approximation for a measure associated with the criterial attribute for voicing in fricative consonants. The selection of this criterion was based on both theoretical and experimental considerations (for detailed discussion cf. Stevens *et al.*, 1992). Briefly, because there is a greater pressure drop across the glottis in the production of voiced fricatives than in the production of voicing in an open vocal

(a) speak [sa]



(b) big [za]

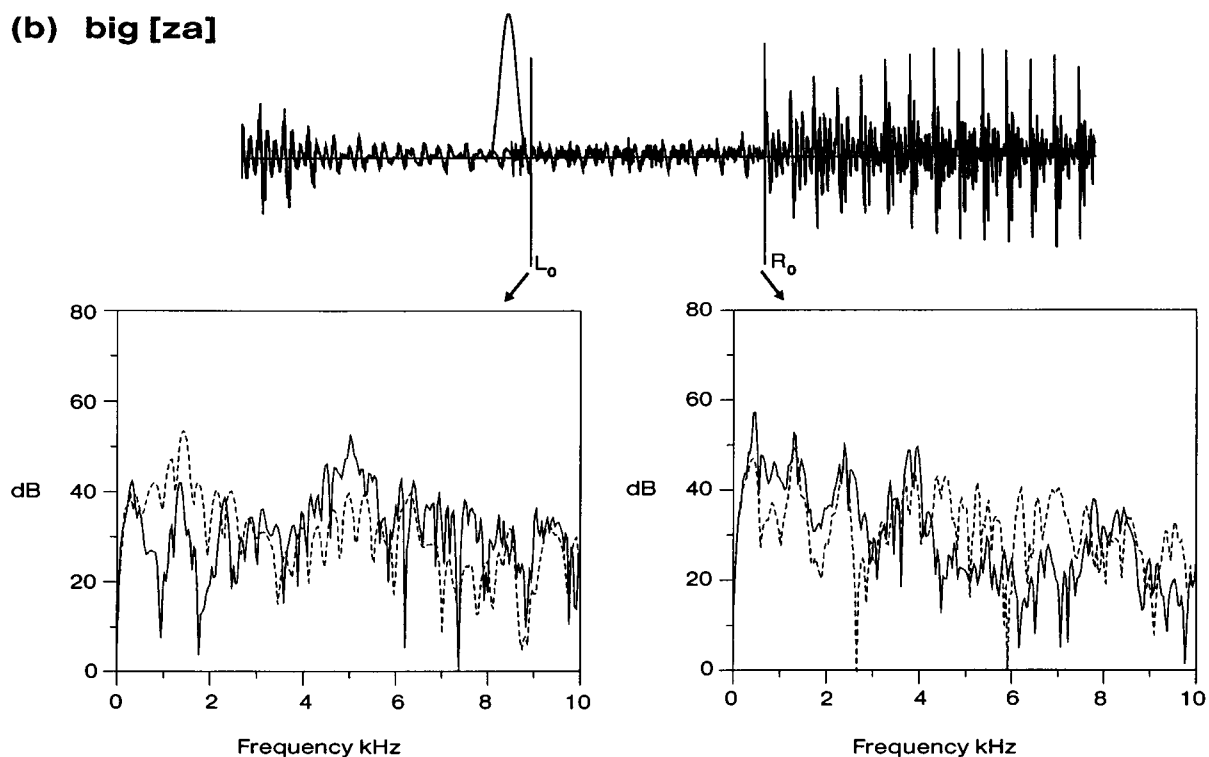


FIG. 1. Examples of cursor placement identifying the fricative noise for two example utterances: "please speak [sa] again" [top panel (a)] and "say a big [za] again" [bottom panel (b)]. The waveform is presented with the left cursor (L_0) placed at the onset of the fricative noise and the right cursor (R_0) placed at the end of the fricative noise. A 15-ms full-Hamming window is shown placed immediately preceding the L_0 cursor. The DFT spectra for the 15 ms immediately preceding (represented by the dotted lines) and following (represented by the solid lines) each cursor placement is shown immediately below the waveform display. See text for discussion.

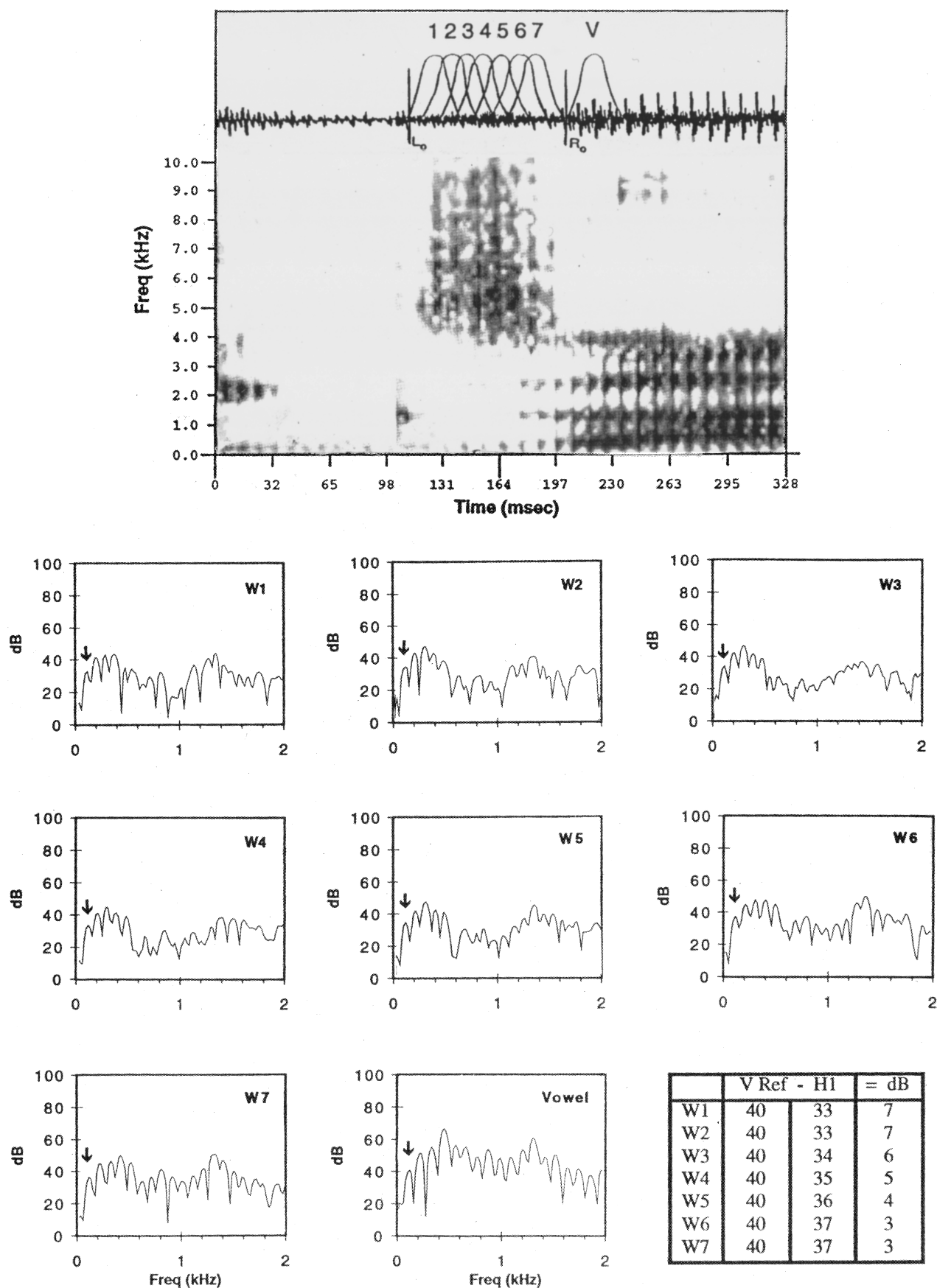


FIG. 2. Illustration of the procedures for applying the voicing metric in fricative consonants. The top panel shows the waveform and spectrogram of the same utterance used in Fig. 1(b) ("big [za]"). The left cursor (L_0) denotes the onset of frication noise and the right cursor (R_0) marks the end of frication. A 30-ms full-Hamming window was advanced in 10-ms steps throughout the frication noise (1–7), and the corresponding amplitude measurements for $H1$ were made (see arrows at W1–W7). The vowel referent (V), defined as the first 30 ms of voicing excluding any frication noise, was determined, and the $H1$ was measured (arrow at dft labeled "Vowel"). The lower right-hand table displays the procedure used for classifying voicing at each window position (see text).

TABLE I. Durations of fricative noise (in ms). For each subject, $N = 30$ tokens per fricative, per condition.

| Condition | Subj. | [f] | [v] | [s] | [z] |
|-------------------|-------------|------------|------------|------------|------------|
| After [k] | 1 | 164 | 107 | 175 | 150 |
| | 2 | 176 | 130 | 203 | 172 |
| | 3 | 148 | 106 | 154 | 114 |
| | 4 | 157 | 106 | 161 | 123 |
| | Mean | 161 | 112 | 173 | 140 |
| After [g] | 1 | 169 | 108 | 186 | 144 |
| | 2 | 178 | 120 | 197 | 155 |
| | 3 | 151 | 104 | 151 | 102 |
| | 4 | 168 | 108 | 162 | 117 |
| | Mean | 167 | 110 | 174 | 130 |
| Isolation | 1 | 226 | 125 | 234 | 169 |
| | 2 | 239 | 148 | 302 | 210 |
| | 3 | 198 | 122 | 210 | 140 |
| | 4 | 193 | 116 | 199 | 149 |
| | Mean | 214 | 128 | 236 | 167 |
| Grand Mean | | 181 | 117 | 194 | 146 |

tract such as in the production of vowels, the amplitude of the first harmonic is calculated to be some 10–15 dB lower for voiced fricatives than that for vowels. Acoustic analyses by Stevens *et al.* (1992) charting the time course of the amplitude of the first harmonic in the vicinity of the vowel-consonant and consonant-vowel boundary are consistent with these calculations. Their perceptual results also show that when the amplitude of the first harmonic of the fricative is some 15 dB lower than that of the vowel, there is a substantial increase in voiceless responses compared to stimuli in which the amplitude of the first harmonic of the fricative is 10 dB down. Moreover, perceptual results suggest that at this amplitude range (0–10 dB) at least 30 ms of voicing is necessary for a stimulus to be categorized consistently as voiced.

II. RESULTS

A. Duration measures

Table I shows the mean durations of the fricative noise interval for voiced and voiceless fricatives. As the results indicate, the mean durations for voiceless fricatives are longer than the mean duration of voiced fricatives for all subjects. The direction of these duration differences replicates the findings of previous studies (e.g., Crystal and House, 1988; and Stevens *et al.*, 1992), although the absolute duration measures obtained in this study were somewhat different from the others, presumably because of the different context conditions across studies. In addition, the duration measures in the current study show an effect for place of articulation similar to that noted in Baum and Blumstein (1987) and Stevens *et al.* (1992). That is, the alveolar fricatives [s] and [z] tend to have longer durations overall than the labiodental fricatives [f] and [v].

Also shown in Table I are consistent differences for all subjects in duration between the context and the isolation conditions. However, further analysis reveals that some asymmetry exists in these duration differences. That is, for

TABLE II. Categorization of voiced and voiceless fricatives by 30-ms metric. Areas of voicing in targets were established by measuring the first harmonic relative to the following vowel (see text). Percentages are based on a total N of 360 utterances per subject, with 60 produced per condition.

| | | Fricatives correctly categorized | | | |
|---------|-----------|-------------------------------------|-----------|-------------------|-----------|
| | | Voiceless targets | | Voiced targets | |
| | | Cond | % | Cond | % |
| Subject | 1 | after [k] | 100 | after [k] | 78.4 |
| | | after [g] | 100 | after [g] | 100 |
| | | isolation | 100 | isolation | 90.0 |
| | | Mean | 100 | Mean | 89.5 |
| | | 2 | after [k] | 100 | after [k] |
| | after [g] | | 100 | after [g] | 86.6 |
| | isolation | | 100 | isolation | 81.7 |
| | Mean | | 100 | Mean | 77.2 |
| | 3 | | after [k] | 100 | after [k] |
| | | after [g] | 100 | after [g] | 98.3 |
| | | isolation | 100 | isolation | 94.9 |
| | | Mean | 100 | Mean | 95.0 |
| | | 4 | after [k] | 98.3 | after [k] |
| | after [g] | | 100 | after [g] | 91.7 |
| | isolation | | 100 | isolation | 73.3 |
| | Mean | | 99.4 | Mean | 82.2 |
| | \bar{X} | | after [k] | 99.6 | after [k] |
| | | after [g] | 100 | after [g] | 94.2 |
| | | isolation | 100 | isolation | 85.0 |
| | | Mean | 99.4 | Mean | 82.2 |
| | | Grand Mean | 99.9 | Mean | 86.0 |

all subjects, the duration differences between the context and the isolation conditions are greater for voiceless fricatives than for voiced fricatives. For example, when all subjects are taken together, the average duration of [f] in isolation is 23% longer than of [f] in context, while the average duration of [v] in isolation is only 13% longer than that of [v] in context. Similarly, the average duration of [s] in isolation is 26% longer than that of [s] in context, while the average duration of [z] in isolation is only 19% greater than that of [z] in context.

B. Voicing measure

1. Categorization of voicing in fricative consonants

The metric for voicing in fricative consonants was applied to all stimuli in all conditions. Table II shows the results. The metric correctly categorized voiceless fricative targets as “voiceless” 99.9% of the time, and correctly categorized voiced fricative targets as “voiced” 86% of the time. Overall, then, the metric correctly categorized 93% of voiced and voiceless fricative targets.

For the voiceless fricatives, only one token for one subject was incorrectly categorized by the metric. In this case, the stimulus token had 30 ms of voicing present at the be-

TABLE III. Categories of different patterns of voicing.

| | |
|-----------------------|--|
| 1. Voicing throughout | Continuous voicing throughout the frication interval. |
| 2. Leading edge | 30 ms of voicing at onset of fricative noise. |
| 3. Leading edge plus | More than 30 ms of voicing at the onset of fricative noise. |
| 4. Trailing edge | 30 ms of voicing at the end of the fricative noise. |
| 5. Trailing edge plus | More than 30 ms of voicing at the end of the fricative noise. |
| 6. Other voicing | No voicing in voiced targets; intermittent voicing but not contiguous. |

ginning of the fricative (although this token was perceived as voiceless by the experimenter). For the voiced fricatives, the metric performed better in some phonetic contexts than in others. For all of the subjects taken together, 94.2% of the voiced targets were correctly identified by the metric when they followed the voiced velar stop [g]. After the voiceless velar stop [k], however, only 78.8% of the voiced targets were correctly identified. The metric's performance on fricatives in isolation fell in between these two context conditions, with 85% of the voiced targets being correctly identified.

The data were further analyzed to determine whether there were any differences in the classification of voicing as a function of place of articulation. This analysis focused on the voiced fricatives only since only one voiceless token was miscategorized. Results can be summarized as follows: In the isolation condition, the average number of misclassifications across the four subjects was four for [v] and three for [z]; in the voiceless context, there were six errors for [v] and four errors for [z]; and in the voiced context, there was one

error for [v] and three errors for [z]. Thus, there were no systematic differences between [v] and [z].

2. Patterns of voicing

Although the metric provides a measure of the presence or absence of voicing in fricative utterances, it does not provide a means of exploring the influence of neighboring voiced or voiceless stop consonants on the voicing characteristics of fricatives. That is, even though fricative consonants may be correctly categorized in terms of the metric for voicing, the implementation of voicing could be affected by phonetic context. Such an influence would be realized in terms of where in the fricative noise voicing emerged. Thus, voicing lead (i.e., voicing beginning at the onset of the fricative noise) should be more frequent for voiced fricatives when preceded by voiced stops than by voiceless stops, and voicing lag (i.e., voicing beginning after the onset of the fricative noise) should be more frequent for voiced fricatives when preceded by voiceless stop consonants. To explore this question, we investigated the patterns of voicing in fricative consonants by determining the time course of voicing onset and offset throughout the fricative noise, and compared these resulting patterns in terms of the phonetic context in which they occurred. Six potential patterns of voicing for voiced fricatives were identified. These patterns are summarized in Table III.

All voiced fricative targets which were categorized correctly as voiced by the metric were analyzed in terms of these patterns. Table IV shows the results of this analysis. As Table IV indicates, while there are individual differences in terms of the percentage of utterances falling within the various categories, all subjects show the effects of phonetic context on the production of voiced fricatives. In particular, when a voiced fricative is preceded by a voiced stop consonant and hence when the vocal cords are already vibrating or

TABLE IV. Distribution of voicing patterns in voiced fricatives. Percentages are of voiced fricative targets that were correctly categorized by the 30-ms metric. (Numbers might not add to 100%, due to rounding.)

| SUBJ | Condition | Voicing patterns | | | | | | N Total= 619 |
|------|-----------|-----------------------|-----------------|----------------------|------------------|-----------------------|------------------|-----------------|
| | | Voicing throughout | Leading edge | Leading edge plus | Trailing edge | Trailing edge plus | Other voicing | |
| 1 | after [k] | 43.6 | ... | 2.1 | ... | 53.2 | 2.1 | 47 |
| | after [g] | 65.0 | 3.3 | 23.3 | ... | ... | 8.3 | 60 |
| | isolation | 68.5 | ... | 5.6 | ... | 22.2 | 3.7 | 54 |
| 2 | after [k] | 2.6 | ... | ... | 10.5 | 81.6 | 5.3 | 38 |
| | after [g] | 50.0 | 3.8 | 21.2 | ... | 11.5 | 13.5 | 52 |
| | isolation | 20.4 | ... | 6.1 | ... | 63.3 | 10.2 | 49 |
| 3 | after [k] | 61.8 | ... | 14.5 | ... | 10.9 | 12.7 | 55 |
| | after [g] | 72.9 | 3.4 | 18.6 | ... | ... | 5.1 | 59 |
| | isolation | 63.2 | ... | 14.0 | ... | 8.8 | 14.0 | 57 |
| 4 | after [k] | 40.8 | 2.0 | 16.3 | ... | 32.7 | 8.2 | 49 |
| | after [g] | 89.1 | ... | 3.6 | ... | 1.8 | 5.5 | 55 |
| | isolation | 50.0 | 4.6 | 15.9 | ... | 11.4 | 18.2 | 44 |
| Mean | after [k] | 37.2 | 0.5 | 8.2 | 2.6 | 44.6 | 7.1 | 189 |
| | after [g] | 69.3 | 2.6 | 16.7 | ... | 3.3 | 8.1 | 226 |
| | isolation | 50.5 | 1.2 | 10.4 | ... | 26.4 | 11.5 | 204 |

TABLE V. Distribution of voicing at onset and voicing delay categories.

| | Voicing at onset | Voicing delay |
|-----------|------------------|---------------|
| after [k] | 45.9 | 47.2 |
| after [g] | 88.6 | 3.3 |
| isolation | 62.1 | 26.4 |

are at least in a configuration for glottal vibration, voicing tends to occur at the onset and throughout all or part of the fricative noise. For example, for all four subjects, the “voicing throughout” pattern occurs more frequently following the voiced velar stop [g], as opposed to the voiceless velar stop [k]. Similarly, for three of the four subjects (S1, S2, S3), the “leading edge” pattern (30 ms of voicing at the beginning of frication) and the “leading edge plus” pattern (more than 30 ms of voicing at the beginning of the fricative) occur more often after the voiced velar stop [g] than after the voiceless velar stop [k].

When a voiced fricative follows the voiceless velar stop [k], and hence where the vocal cords are presumably in a configuration that inhibits vocal cord vibration, there is a short lag in the initiation of voicing in the fricative noise. For all subjects, the “trailing edge plus” pattern (more than 30 ms of voicing at the end of the fricative) occurs more often after the voiceless velar stop [k]. The “trailing edge” pattern (30 ms of voicing at the end of the fricative) occurs infrequently, but always after [k]. For all subjects, the “voicing throughout” pattern occurs less often following the voiceless velar stop [k].

The distribution of voicing patterns for voiced fricatives produced in isolation resembles the distribution of those following the voiceless velar [k]. This pattern is not unexpected, as in both cases, the vocal cords must be set into motion or closed to produce the voiced fricative. In general, the “voicing throughout” pattern occurs less often in isolation than after the voiced velar stop [g]; the same is true for the “trailing edge plus” pattern of voicing. Voiced fricatives in the isolation condition were produced in the absence of the coarticulation effects of a preceding voiced or voiceless consonant. Thus, individual differences in voicing pattern distribution among the speakers may be more likely to emerge reflecting varying strategies for setting the vocal cord apparatus into motion. For example, one subject tended to precede both voiced and voiceless fricatives produced in isolation with a glottal stop.

Voicing patterns in the “other” category include intermittent types of voicing. These patterns are randomly distributed among the four subjects, with no discernible effect across condition (context versus isolation) or environment (following a voiceless versus a voiced velar stop).

The various categories of voicing can be collapsed further into two broad categories, one characterized by voicing from the onset of the fricative noise, and the other characterized by a delay in voicing onset relative to the onset of the fricative noise. The first category, called *voicing at onset*, includes voicing throughout, leading edge, and leading edge plus, and the second category, called *voicing delay*, includes trailing edge and trailing edge plus. The influence of pho-

TABLE VI. Distribution of voicing at the leading edge in voiceless fricative targets. Percentages are based on the number of tokens wherein voicing occurs at the leading edge, out of the total number of voiceless fricative targets that contain some voicing (N).

| | Condition | % voicing at leading edge | N |
|-----------|-----------|---------------------------|-----|
| Subject 1 | after [k] | ... | 0 |
| | after [g] | ... | 0 |
| | isolation | ... | 0 |
| 2 | after [k] | 0 | 1 |
| | after [g] | 100.0 | 5 |
| | isolation | 0 | 1 |
| 3 | after [k] | ... | 0 |
| | after [g] | 100.0 | 2 |
| | isolation | ... | 0 |
| 4 | after [k] | 50.0 | 2 |
| | after [g] | 85.7 | 7 |
| | isolation | ... | 0 |

netic context can be seen by examining the extent to which voicing at onset emerges as the predominant pattern in the context of a voiced stop and voicing delay emerges in the context of a voiceless stop. Table V shows the results. It is clear from this analysis that the influence of the voiced stop on the pattern of voicing is greater than that of the voiceless stop. In particular, all subjects show a predominance of the voicing at onset pattern in the context of a preceding voiced stop. In contrast, while the preceding voiceless stop clearly influences the extent to which voicing delay occurs for fricatives, it is the predominant pattern for only two of the four subjects (S1 and S2). In isolation, the preferred pattern of voicing for fricatives for three of the four subjects (S1, S3, S4) is voicing at onset.

With respect to the voiceless fricative tokens, it is also possible to explore whether phonetic context influences the patterns of voicing. As discussed above, 99% of the voiceless fricative targets were categorized correctly as voiceless. Nonetheless, the effect of coarticulation would predict that voicing, if it occurs in a voiceless fricative target, would be expected at the leading edge after a voiced stop. To determine whether this is the case, the correctly categorized voiceless fricative tokens were analyzed to determine whether any measured voicing (i.e., 10 or 20 ms of voicing by the metric) occurred in the fricative noise interval, and where in the interval it occurred. As Table VI indicates, the pattern of results shows that the preceding voiced and voiceless stop influenced the production of voiceless fricatives as well. Although the preponderance of voiceless fricatives had no voicing in the fricative noise, sixteen of the 18 utterances which contained some voicing in the voiceless fricative noise interval had voicing lead (i.e., at the onset of the fricative noise). Of these 16 utterances, 14 (or 88%) occurred following the voiced velar stop [g].

To determine whether there were any differences in the patterns of voicing as a function of place of articulation of the fricative consonant, the data were further analyzed and classified into three categories depending on whether there was voicing throughout the duration of the fricative noise, or

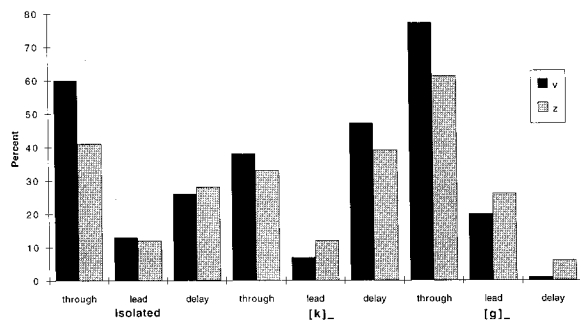


FIG. 3. Patterns of voicing as a function of place of articulation of the fricative consonant ([v z]) across the three speaking conditions (isolated, following [k], and following [g]).

whether there was only voicing lead or voicing delay. Figure 3 shows the results. For both [v] and [z], the pattern of voicing was influenced similarly by the context condition. When the voiced fricative was preceded by the voiceless velar stop, there was a decrease in the occurrence of the pattern of voicing throughout and an increase of voicing delay; when the voiced fricative was preceded by a voiced velar stop consonant, there was an increase in the occurrence of the pattern of voicing throughout and voicing lead. The only difference that emerged between [v] and [z] was a greater preponderance of voicing throughout for the labiodental fricatives compared to the alveolar fricatives, particularly in the isolation and in the voiced context conditions.

C. Errors in categorization of voicing in fricatives

The previous analyses explored the patterns of voicing that emerged for voiced and voiceless tokens correctly classified by the 30-ms metric. In order to determine whether the so-called classification errors emerged as a function of the voicing of the preceding stop consonant, we conducted analyses of those targets which failed to be correctly classified by the metric. (This analysis was restricted to voiced fricative targets since only one voiceless target was misclassified by the metric.) Voiced targets failed to be categorized as voiced for three reasons: (1) Voicing was present at amplitudes of 10 dB or less but there was fewer than 30 ms of contiguous voicing at either edge of the fricative; (2) there was glottal excitation of 30 ms or more at either edge of the fricative noise, but it was not of sufficient amplitude to be considered voiced by our measure; (3) there was no voicing or glottal excitation present in the fricative noise interval. The three types of categorization errors taken together represent a kind of “continuum” in the strength of glottal excitation. The results are summarized in Table VII.

As Table VII shows, the distribution of the types of voicing “errors” appears to be variable across speakers, except for the category in which no voicing occurred in the fricative noise. For three of the four subjects who produced this type of error, it occurred most frequently after the voiceless velar stop [k], followed in frequency of occurrence by the isolation condition. This type of error never occurred after the voiced stop [g]. The greatest total number of errors occurred after the voiceless velar stop, followed again by the isolation condition. Only 11% of these “errors” occurred in

TABLE VII. Analysis of voiced fricative targets incorrectly categorized by the metric as voiceless. See text for an explanation of the error categories. Percentages are based on a total N of “errors” (i.e., voiced targets which were not categorized as “voiced”). (Numbers might not add to 100%, due to rounding.)

| | | | Duration of voicing < 30 ms | Amplitude of voicing < 10 dB | No voicing | N |
|---------|-----------|-----------|-----------------------------------|------------------------------------|------------|----|
| Subject | Condition | | | | | |
| 1 | after [k] | 15.4 | 38.5 | 46.2 | 13 | |
| | after [g] | 0 | 0 | 0 | 0 | |
| | isolation | 50.0 | 33.3 | 16.7 | 6 | |
| | 2 | after [k] | 36.4 | 40.9 | 22.7 | 22 |
| | | after [g] | 0 | 100.0 | 0 | 8 |
| | | isolation | 54.5 | 27.3 | 18.2 | 11 |
| | 3 | after [k] | 20.0 | 80.0 | 0 | 5 |
| | | after [g] | 100.0 | 0 | 0 | 1 |
| | | isolation | 33.3 | 66.7 | 0 | 3 |
| | 4 | after [k] | 45.5 | 36.4 | 18.2 | 11 |
| | | after [g] | 40.0 | 60.0 | 0 | 5 |
| | | isolation | 12.5 | 81.3 | 6.3 | 16 |

the context of the voiced stop [g]. Thus, similar to the analysis of the patterns of voicing in correctly categorized voiced and voiceless fricative productions, there was an influence of the voicing of preceding stop consonants on voicing in the fricative consonants.

III. DISCUSSION

This study investigated the acoustic characteristics of voicing in fricative consonants. The results indicate that, although the voicing characteristics of a preceding voiced or voiceless velar stop influence the patterns of voicing in initial fricative consonants, the acoustic properties corresponding to the phonetic category of voicing remain intact. While there may be influences of context on the realization of voicing, those influences affect only the patterns of voicing and not the criterial attribute of 30 ms of glottal excitation at the acoustic boundaries of the fricative noise. In the current study, the criterial attribute reliably categorizes the majority (93%) of the stimuli.

Several points are worth making concerning the selection of the acoustic parameters used to define the criterial attribute and the overall success of the attribute in classifying correctly the stimuli. Although the choice of a 10-dB cutoff and a 30-ms duration was based on acoustic and perceptual evidence as well as theoretical considerations, the actual criteria selected were in some sense arbitrary. The acoustic parameters selected served as a first approximation for determining whether an acoustic measure could classify voicing in fricative consonants and also whether such a measure could provide a means for investigating the patterns of voicing that might emerge across different phonetic contexts. The establishment of the particular criteria used indicated that indeed an acoustic property focussing on the presence/absence of glottal excitation could be identified, that this acoustic property could characterize voicing across a number of sources of variability (speakers, vowel contexts, and phonetic environments), and that the property could also provide

insight into the time course of the implementation of voicing in fricative consonants as it was influenced by the voicing of a preceding stop consonant. Changing the threshold of the acoustic criterion (either in the amplitude or time domain) would have changed the results some but not substantially. Similarly, changing the laboratory conditions might have had some effect on the obtained results, although again, we would propose not in a substantial way. The measures used in this study were made on speech samples collected under controlled conditions in a laboratory setting, and presumably the conditions were in this sense optimal. Nonetheless, we would expect that similar patterns of results would emerge if the speech samples were collected in more natural settings, although the success rate of the measure would presumably not be as high.

As to the particular acoustic property identified as the criterial attribute for voicing in fricative consonants, we do not believe that vowel length could serve a similar role. As we have discussed earlier, a criterial attribute, by definition, is an acoustic property that serves as a defining parameter for a particular phonetic dimension. As such, it must serve as a stable property across numerous sources of variability, and serve a role in perception. Although vowel length may play an important role in the perception of voicing in final fricative consonants, it does not serve such a role in initial fricative consonants (the phonetic environment explored in this study). As shown elsewhere, considerable overlap has been demonstrated in measures of the production of the duration of voicing in initial position (Baum and Blumstein, 1987; Crystal and House, 1988). Likewise, perception experiments (Jongman, 1989) have shown that in initial position, the duration of a voiceless fricative can be reduced to as little as 20 ms, and listeners will still perceive the stimuli as voiceless. Although this study did not explore the perceptual role of the proposed criterial attribute, Stevens *et al.* (1992) did show that the presence/absence of about 30 ms of voicing (defined similarly as in this paper) played a critical role in the perception of voicing in fricative consonants in both initial and final position.

In the context of the overall stability of the criterial attribute, the results of this study show that the patterns of implementation of voicing vary as a function of the phonetic context. In particular, not surprisingly, the voicing characteristics of a preceding stop influence where in the fricative noise voicing emerges. When a voiced consonant precedes a voiced fricative, voicing in the fricative noise is more likely to occur at the onset of the fricative noise. In contrast, when a voiceless consonant precedes the fricative, voicing onset in the fricative noise is more likely to be delayed. The correct classification of voicing by the metric is also influenced to some extent by the voicing characteristics of the preceding stop consonant. The acoustic measure made a greater number of classification errors for voiced fricatives when preceded by a voiceless stop. In contrast, voiceless fricatives were less likely to be influenced by the voicing of the preceding stop consonant both in terms of correct classification and the patterns of voicing. Nonetheless, there was still an influence of context. There were some cases of glottal excitation occurring at the onset of the fricative noise in voiceless fricatives,

especially when the fricative noise was preceded by a voiced stop consonant.

Having said this, although contextual influences emerged, they did not necessarily occur in the majority of the utterances, and there was variation among the four speakers. In most cases, the preferred pattern of voicing in fricative consonants across speakers and phonetic environments was the presence of voicing throughout the fricative noise for voiced fricatives and the absence of such voicing for voiceless fricatives. Even in those examples of voiced fricatives where voicing onset was delayed, analysis of the individual tokens revealed that there was usually glottal excitation at the onset of the noise but it was not sufficiently strong to meet the requirements of voicing by the metric. In many cases, the glottal pulsing failed to meet the metric by only 1 or 2 dB. Moreover, the interval in which voicing delay occurred was usually no more than 10 or 20 ms after the onset of the fricative noise. Thus, glottal excitation was usually present at the onset of voiced fricatives regardless of the phonetic context, but the voicing of the preceding stop consonant influenced the strength of that voicing at fricative onset. Even in those voiced fricative productions which were classified incorrectly by the metric as voiceless, glottal excitation was present in nearly all of the utterances and typically throughout the duration of the fricative noise. The amplitude of that excitation, however, was not strong enough for the fricative to be classified as voiced by our analysis procedures.

That voiced fricatives were more likely influenced by phonetic context than voiceless consonants in terms of both classification errors and patterns of voicing is not surprising. Once the glottal area is sufficiently large in the production of a voiceless fricative, there is considerable latitude in adjusting the area of the glottal opening and of the supraglottal constriction while still maintaining voicelessness and strong frication noise. In contrast, for a voiced fricative, the glottal and supraglottal constrictions must maintain a proper balance. Increasing or decreasing the area of one of the constrictions would upset that balance, and either sustained voicing or sustained frication would be jeopardized. As such, it may be harder to sustain voicing than to sustain voicelessness in fricative consonants. Interestingly, there is a prevalence of voiceless fricatives over voiced fricatives in the languages of the world (Maddieson, 1984), perhaps reflecting the greater difficulty in producing voiced compared to voiceless fricatives.

Both classification errors by the acoustic measure and the patterns of voicing that emerged showed the greatest effects when the two contiguous consonants had mixed voicing. For all of the subjects, the majority of voiced fricative targets that failed to be categorized as voiced occurred following voiceless velar stops. Very few errors occurred when a voiced fricative target followed a voiced velar stop. Similarly, for voiceless fricative targets following a voiced velar stop, voicing sometimes continued into the voiceless fricative. There were virtually no instances of glottal excitation at the onset of voiceless fricatives when preceded by a voiceless velar stop. This pattern for both voiced and voiceless fricatives was also seen in the analysis of patterns of voicing.

For voiced fricative targets following a voiceless velar stop, voicing was less likely to continue throughout the fricative than it was for voiced fricative targets preceded by a voiced stop. In addition, voicing occurred at the trailing edge of a fricative following a voiceless stop (i.e., at the fricative–vowel boundary) far more frequently than for voiced fricatives following a voiced stop. Thus, it seems to be easier to sustain voicing and/or voicelessness across segmental boundaries than it is to make the rapid adjustments coordinating the articulatory supraglottal and glottal gestures required for the production of contiguous consonants with mixed voicing.

Stevens *et al.* (1992) showed an asymmetry in the influence of voicing in those two fricative sequences which had mixed voicing. In particular, they found less of an influence of voiced fricatives on voiceless fricatives than the reverse. Thus, voiceless fricatives tended to inhibit glottal vibration in an adjacent voiced fricative. We found similar results. In particular, a preceding voiceless stop tended to inhibit the onset of glottal vibration in a following voiced fricative. Only 79% of voiced fricatives had voicing lead in the environment of [k] compared to 94% in the environment of [g]. This pattern of results is also reflected in the analysis of classification errors, where the greatest number of classification errors for voiced fricatives occurred following a [k]. Nonetheless, overall, the voiced stop [g] had a more consistent influence than the voiceless stop [k] on voiced fricatives. Voicing lead occurred in 89% of the utterances after [g], whereas after [k], the patterns of voicing were evenly distributed across the voicing lag and voicing lead categories.

It has been suggested that the patterns of voicing are affected by the place of articulation of the fricative consonant. In particular, Haggard (1978) showed a greater occurrence of devoicing in the alveolar fricative [z] than in the labiodental fricative [v]. We also found that subjects were more likely to show voicing throughout the duration of the fricative noise in the production of [v] than in the production of [z]. Aside from that difference, analyses of both the classification of voicing by the metric and the patterns of voicing showed few differences between the fricative consonants. The fact that there was less voicing throughout the duration of the fricative noise in the case of [z] could reflect the difficulty of maintaining voicing over the longer durations inherent in the alveolar fricative compared to the labiodental fricative, and it could also reflect a difference in vocal tract compliance between [z] and [v].

Overall, the results of the current study show better classification scores by the metric for fricative voicing than those obtained by Stevens *et al.* (1992). As discussed earlier, one of the reasons for the poorer classification scores in the Stevens *et al.* (1992) study might be attributable to the difficulty inherent in the control of voicing across contiguous fricatives. Another consideration is the difficulty noted by Stevens *et al.* in determining the acoustic “boundary” between the two fricative consonants. The determination of the boundary is crucial for the application of the voicing metric. A third consideration is related to the nature of the stimulus materials used. Subjects were asked to read phrases with fricatives appearing in various environments, some of which

were similar to tongue twisters (e.g., “his fav sips it”). These types of phrases may have induced such speech production errors. Mowrey and Mackay (1990) have shown using EMG evidence that the majority of tongue-twister production errors are “subfeatural” errors; that is, they fall within a range between “full errors” and “non-errors.” These subfeatural errors, although detectable by EMG, are unnoticed by speakers and listeners. It is possible that a number of the classification errors in the Stevens *et al.* (1992) study reflected such articulatory errors by the subjects.

The current study investigated the effects of left to right or carryover coarticulation on voicing in fricative consonants. The extent of such coarticulatory effects has been used to investigate the size and nature of the organizational units in speech production. It has been generally assumed that left to right coarticulation reflects mechanico-inertial forces in the process of the implementation of articulatory gestures (Daniloff and Hammarberg, 1973; Bell-Berti *et al.*, 1995; Kent and Minifie, 1977). The results of this study are consistent with the view that such carryover effects are short-lived, influencing the first 10’s of ms of the following segment, and that these effects arise from the dynamics of gestural overlap in the production of a given segment, in this case presumably the glottal opening (Fowler and Saltzman, 1993).

In sum, the current study showed that carryover articulation influences the production of voicing in fricative consonants. These coarticulatory effects are short-lived, however, and affect only the patterns of voicing that emerge in the fricative noise, not the acoustic property corresponding to the phonetic feature for voicing. Thus, there is both variability and stability in the implementation of the phonetic category of voicing. There is variability in the duration of glottal excitation during the frication noise and in the patterns of voicing, and there is stability in the acoustic property or criterial attribute corresponding to the phonetic feature of voicing. This stability remains present across sources of variability contributed by different speakers, vowel contexts, and phonetic environments.

ACKNOWLEDGMENTS

This research was supported in part by NIH Grant No. DC00142 to Brown University. Many thanks to Martha Burton and Kenneth N. Stevens for their comments on an earlier draft of this paper.

¹We recorded eight blocks of data, but only used six for acoustic analysis. The other two blocks were available in the event that individual tokens needed to be replaced.

²The number of glottal pulses that fell within the 30-ms window for each speaker was three pulses for the males and five pulses for the females.

Baum, S., and Blumstein, S. (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.

Bell-Berti, F., Krakow, R., Gelfer, C., and Boyce, S. E. (1995). “Anticipatory and carryover implications for models of speech production” in *Producing Speech: Contemporary Issues*, edited by F. Bell-Berti and L. J. Raphael (American Institute of Physics, New York).

- Blumstein, S. E., and Stevens, K. N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," *J. Acoust. Soc. Am.* **66**, 1101–1017.
- Crystal, T., and House, A. (1988). "A note on the durations of fricatives in American English," *J. Acoust. Soc. Am.* **84**, 1932–1935.
- Daniloff, R. G., and Hammarberg, R. E. (1973). "On defining coarticulation," *J. Phonetics* **1**, 239–248.
- Denes, P. (1955). "Effect of duration on the perception of voicing," *J. Acoust. Soc. Am.* **27**, 761–764.
- Fowler, C. A., and Saltzman, E. (1993). "Coordination and coarticulation in speech production," *Language and Speech* **36**, 171–195.
- Haggard, M. (1978). "The devoicing of voiced fricatives," *J. Phonetics* **6**, 95–102.
- Jongman, A. (1989). "Duration of friction noise required for identification of English fricatives," *J. Acoust. Soc. Am.* **85**, 1718–1725.
- Kent, R., and Minifie, F. D. (1977). "Coarticulation in recent speech production models," *J. Phonetics* **5**, 115–134.
- Lehiste, I. (1970). *Suprasegmentals* (MIT, Cambridge).
- Maddieson, I. (1984). *Patterns of Sound* (Cambridge U. P., London).
- Mowrey, R. A., and MacKay, I. R. A. (1990). "Phonological primitives: Electromyographic speech error evidence," *J. Acoust. Soc. Am.* **88**, 1299–1312.
- Slis, I. H., and Cohen, A. (1969). "On the complex regulating the voiced–voiceless distinction," *Language and Speech* **12**, 80–102.
- Stevens, K. N., and Blumstein, S. E. (1981). "The search for invariant acoustic correlates of phonetic features," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Earlbaum, Hillsdale, NJ).
- Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (1992). "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters," *J. Acoust. Soc. Am.* **91**, 2979–3000.
- Utman, J. A., and Blumstein, S. E. (1994). "The influence of language on the acoustic properties of phonetic features: A study of the feature [strident] in Ewe and English," *Phonetica* **51**, 221–238.

Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants

Margaret W. Skinner, Laura K. Holden, and Timothy A. Holden

*Department of Otolaryngology—Head & Neck Surgery, Washington University School of Medicine,
517 South Euclid Avenue, Campus Box 8115, St. Louis, Missouri 63110*

Marilyn E. Demorest

*Department of Psychology, University of Maryland Baltimore County, 5401 Wilkens Avenue,
Baltimore, Maryland 21228*

Marios S. Fourakis

*Department of Speech and Hearing Science, The Ohio State University, 110 Pressey Hall,
1070 Carmack Road, Columbus, Ohio 43210*

(Received 6 March 1996; revised 10 January 1997; accepted 28 January 1997)

Ten postlinguistically deaf adults who used the Nucleus Cochlear Implant System and SPEAK speech coding strategy responded to vowels, consonants, words, and sentences presented sound-only at 70, 60, and 50 dB sound-pressure level. Highest group mean scores were at a raised-to-loud level of 70 dB for consonants (73%), words (44%), and sentences (87%); the highest score for vowels (70%) was at a conversational level of 60 dB. Lowest group mean scores were at a soft level of 50 dB for vowels (56%), consonants (47%), words (10%), and sentences (29%); all except subject 7 had some open-set speech recognition at this level. For the conversational level (60 dB), group mean scores for sentences and words were 72% and 29%, respectively. With this performance and sound-pressure level, it was observed that these subjects communicated successfully in a variety of listening situations. Given these subjects' speech recognition scores at 60 dB and the fact that 70 dB does not simulate the vocal effort used in everyday speaking situations, it is suggested that cochlear implant candidates and implantees be evaluated with speech tests presented at 60 dB instead of the customary 70 dB sound-pressure level to simulate benefit provided by implants in everyday life. Analysis of individuals' scores at the three levels for the four speech materials revealed different patterns of speech recognition among subjects (e.g., subjects 1 and 5). Future research on the relation between stimuli, sound processing, and subjects' responses associated with these different patterns may provide guidelines to select parameter values with which to map incoming sound onto an individual's electrical dynamic range between threshold and maximum acceptable loudness level to improve speech recognition. © 1997 Acoustical Society of America. [S0001-4966(97)00306-8]

PACS numbers: 43.71.Es, 43.66.Ts [WS]

INTRODUCTION

In their classic study of speech levels in everyday life, Pearsons *et al.* (1976) measured the spectra and overall level of speech spoken by males, females, and children at five vocal efforts 1 m from the talkers. One meter approximates the average distance between two people when they carry on a conversation. Average overall levels for casual, normal, raised, loud, and shout vocal efforts were 56, 60, 66, 74, and 83 dB sound-pressure level, respectively. In this same study, speech levels measured for face-to-face conversation averaged 55 dB when background noise was below 48 dB, and as noise increased to 70 dB, speech levels were raised to 70 dB sound-pressure level. These measurements were made with normal-hearing people. When someone speaks to a person with a severe-to-profound hearing impairment who has difficulty understanding what is said, 65–70 dB sound-pressure level may be used to facilitate communication. However, the vocal effort to speak at these levels is almost impossible to maintain for more than a short time. Consequently, people

with this degree of hearing impairment who use hearing aids often are unable to understand general conversation spoken at 1 m.

During the 1970s and 1980s, a number of procedures were developed for prescribing the real-ear gain of linear hearing aids based on speech spectra for an overall level of 65–70 dB sound-pressure level (e.g., Byrne and Dillon, 1986; Byrne and Tonisson, 1976; Cox, 1983; Pascoe, 1978). Studies of hearing-aid gain values used in everyday life supported the use of this range of overall levels (Reid *et al.*, 1977; Walden *et al.*, 1977). It is assumed that the choice of hearing-aid gain values in these studies was strongly affected by the distortion associated with momentary input levels of over 70 dB in the class A amplifiers and relative lack of attention to setting the maximum output so that amplified sound did not exceed an individual's uncomfortable loudness level. That is, individuals probably set hearing-aid gain lower than might have been optimal so that amplified sound was not distorted nor uncomfortably loud. With class D am-

TABLE I. Subject biographical information.

| Subject | Sex | Etiology | Years of profound deafness | Age at implantation (yr) | Number of rings inserted | Months of SPEAK use |
|---------|-----|---|----------------------------|--------------------------|--------------------------|---------------------|
| 1 | M | Otosclerosis | 17 | 60 | 24 | 3 |
| 2 | M | Autoimmune disease | 4 | 44 | 30 | 15 |
| 3 | M | Otosclerosis | 10 | 61 | 32 | 15 |
| 4 | M | Acoustic trauma; progressive loss | 4 | 61 | 32 | 15 |
| 5 | F | Progressive loss; maternal rubella | 22 | 45 | 32 | 15 |
| 6 | F | Progressive loss; meningitis | 9 | 26 | 24 | 15 |
| 7 | M | Vascular; diabetes | 2 | 76 | 22 | 3 |
| 8 | F | Unknown | 15 | 70 | 32 | 2 |
| 9 | F | Progressive loss; prematurity | 10 | 39 | 24 | 2 |
| 10 | M | Hereditary; congenital (LE); progressive (RE) | 10 | 61 | 32 | 4 |

plifiers and full-dynamic range compression available in present-day hearing aids as well as the use of more sophisticated fitting procedures, hearing-aid gain can be set so conversational speech levels (e.g., 60 dB sound-pressure level) can be heard with relative ease by those with mild or moderate hearing impairment. In contrast, those with severe-to-profound impairment cannot understand most of what is said at conversational speech levels because hearing-aid gain must be limited below that which causes acoustic feedback.

When testing protocols were developed to evaluate candidates for cochlear implantation and to determine their ability to recognize speech by sound alone after implantation, 65 or 70 dB sound-pressure level was used (e.g., Owens *et al.*, 1985; Dowell *et al.*, 1987). This choice probably was based on the procedures developed for fitting linear hearing aids described above. As cochlear implantees' speech recognition improved with upgrades in speech coding strategies of the Nucleus 22 Channel Cochlear Implant System (e.g., Dowell *et al.*, 1987; Skinner *et al.*, 1991; Skinner *et al.*, 1994), adults and children have been able to understand speech spoken during conversational speech with relative ease and at distances greater than one meter. For example, adults report that they can recognize words spoken from the next room at distances of 3 m or more. At this distance, speech levels at the implant microphone are at least 10 dB lower than when spoken at the same vocal effort at 1 m. In addition, parents and teachers have observed profoundly deaf children who wore hearing aids and others who used cochlear implants in the same classroom. They noted that the implanted children were able to recognize softer sounds, including speech, at greater distances than those with hearing aids (Skinner, 1994). This ability to hear sound at a distance allows deaf children with implants to attach meaning to sound in spontaneous learning situations and not just during close-range, face-to-face instruction.

There is substantial information about implantees' recognition of speech materials presented at 70 dB sound-pressure level. Except for a recent study by Müller-Deile

et al. (1995) of adults' recognition of German speech materials using the Nucleus MPEAK speech coding strategy (Skinner *et al.*, 1991), there is a paucity of data at levels lower than 70 dB sound-pressure level. At the present time, audiological criteria for implantation are based on aided speech recognition of test materials presented at the raised-to-loud vocal effort of 70 dB sound-pressure level. To develop criteria that more closely approximate the benefit provided by cochlear implants in everyday life, it is important to know how well implantees recognize speech at conversational as well as soft levels. For this reason, the present study was designed to evaluate adult implantees' recognition of vowels, consonants, words, and sentences presented in quiet at 50, 60, and 70 dB sound-pressure level. These levels simulate soft, conversational, and raised-to-loud vocal efforts spoken at 1 m. The soft level also simulates speech spoken at distances of more than 1 m.

I. METHOD

A. Subjects

Ten postlinguistically deaf adults, who use the SPEAK speech coding strategy (Seligman and McDermott, 1995) of the Nucleus 22 Channel Cochlear Implant System, participated. Biographical information about the subjects is given in Table I. The maximum number of electrodes plus supporting rings that can be implanted is 32; if this depth of insertion is achieved in an average length cochlea (approximately 34 mm), active electrodes are positioned near auditory neurons whose characteristic frequencies range from approximately 500 to 4000 Hz. Thirty-two were inserted for five subjects; for four subjects, resistance met during surgery prevented full insertion of all rings; one subject had approximately 5 mm of otosclerotic bone drilled out of the basal turn prior to array insertion. Subjects had used the SPEAK strategy between two and 15 months prior to this study. Previously, all except subject 10 had used the MPEAK strategy. Although these subjects represent a range of speech recogni-

tion abilities, they were chosen for this study because they were using the SPEAK strategy, and they were able to participate.

B. Test materials

American English speech materials that represent three levels of linguistic complexity were used: nonsense syllables, single-syllable words, and sentences. The nonsense syllables included a set of 19 pure vowels, r-colored vowels, and diphthongs (i, I, ε, æ, α, ɔ, ʊ, ʌ, ɪə, ɜ, ʊə, ɒə, ɛə, eɪ, oʊ, aɪ, aʊ, ɔɪ) in a /hVd/ context and a set of 18 consonants (p, b, t, d, k, g, s, z, ʃ, ʒ, tʃ, dʒ, f, v, m, n, l, r) in an /aCa/ context. The vowel set is larger than that used by most investigators. For this study, diphthongs and r-colored vowels were included to determine how well the change in formant frequencies (F_1 and F_2 for diphthongs, and F_3 for r-colored vowels) was tracked by the implant speech processor and perceived by the subjects as compared with the relatively steady-state formants of pure vowels. Furthermore, these large vowel and consonant sets made possible an extensive feature analysis of subjects' responses. Recordings of these vowels and consonants were made by a male talker with midwestern dialect of American English and digitized with 12-bit analog-to-digital conversion. Three repetitions of each vowel were pseudorandomized to form six lists (57 items/list); the same procedure was used for consonants (54 items/list). Six of the 50-word Consonant-Vowel Nucleus-Consonant lists (CNC; Peterson and Lehiste, 1962; lists 3, 5, 6, 7, 9, and 10) were used for testing. Cassette tape recordings of these word lists as well as the City University of New York (CUNY) Topic Sentences (Boothroyd *et al.*, 1988) were made for Cochlear Corporation by a different male talker with midwestern dialect of American English. Twenty-four of the 72 CUNY Sentence lists were used for testing; in each list, there were 12 sentences containing 102 words that were scored. Different word and sentence lists were used for practice prior to testing. Because a normal vocal effort was used in making these materials, the usual change in speech spectrum contour with change in vocal effort was not represented in this experiment. For this reason, the use of three levels only simulated different vocal efforts. To integrate sound levels over entire lists of vowels, consonants, words and sentences, the L_{eq} setting on the sound-level meter was used. This L_{eq} measurement is the same as the long-term, rms sound-pressure level for either one-third octave bands or broadband stimuli. Its measurement required that a second set of recordings be made from the original recordings in which the vowels, consonants, and words were digitally abutted to eliminate silent periods.

Warble tones were used to obtain sound-field thresholds. These were triangular-wave, modulated tones centered at 250, 500, 1000, 1500, 2000, 3000, 4000, and 6000 Hz of the standard bandwidths recommended for use in the sound field by Walker *et al.* (1984).

C. Equipment/test environment

The Nucleus 22-electrode intracochlear implant has been described by Clark *et al.* (1987), and the Spectra 22

speech processor and SPEAK speech coding strategy have been described by Seligman and McDermott (1995).

A PDP-11/23+ computer interfaced to a custom-built audiological console was used to control presentation of the speech materials and warble tones as well as collection and scoring of subjects' responses to vowels and consonants obtained with a touch-display screen (Electro Mechanical Systems, model TID). Warble tones were generated by a multi-function synthesizer (Hewlett-Packard, model 8904A). A cassette player (Tascam, model 112) was used to present the tape-recorded tests. The audio signal was presented through a Urei loudspeaker (model 809). The overall level (sound-pressure level, dB) of the vowels, consonants, words, and sentences was the average of peaks on the slow, rms, linear scale of the sound-level meter (Brüel & Kjaer, model 2230; microphone: Brüel & Kjaer, model 4155). Third-octave-band, L_{eq} measurements (dB) of the vowels, consonants, words, sentences, and warble-tones were made with the same sound-level meter, microphone, and a filter set (Brüel & Kjaer, model 1625). For both sets of measurements, the microphone was placed at the center of where the subject's head would be during testing.

Recorded speech tests and warble tones were presented in a double-walled sound booth (IAC, model 1204-A; 254 × 264 × 198 cm) with the subjects seated at 0° azimuth, 1.5 m in front of the loudspeaker, the center of which was 118 cm above the floor.

D. Procedures

1. Speech processor settings

A data set called a MAP was created for each subject using threshold and maximum acceptable loudness levels for each active electrode for the chosen electrode pairing mode, assignment of frequency boundaries to electrodes, base level and Q level, preamplifier gain, and SPEAK speech coding strategy (for description of these parameters, see Skinner *et al.*, 1991, 1994). Values selected for these parameters were modified until subjects found conversational speech (approximately 60 dB sound-pressure level) in everyday life comfortably loud at the optimum sensitivity control setting (Seligman and Whitford, 1995). The sensitivity control adjusts the preamplifier gain of the microphone input signal to the speech processor. When it is set so that the automatic gain control (that follows the preamplifier in the circuit) is just on the point of operating at peaks of 74 dB sound-pressure level, the largest range of input signals is mapped onto most, if not all, of the electrical range between minimum (threshold) and maximum (maximum acceptable loudness level) stimulation levels specified in each subject's MAP. For this reason, it is called the "optimum" setting. MAP parameters also were modified so that loud sounds were not uncomfortable, soft sounds were audible, and speech was as clear as possible. Because MAP parameters were set according to these criteria, subjects used their processors at the optimum sensitivity settings for most listening situations in everyday life as well as for this study (mean

TABLE II. Features specified for information transmission analysis.

| Vowels | | Features | | | | | |
|--------|--------|-----------------|----------------------|---------------------|-------|----------|-----------|
| Token | Symbol | <i>r</i> -color | Height (<i>F</i> 1) | Front (<i>F</i> 2) | Tense | Duration | Diphthong |
| heed | i | 0 | 0 | 0 | 1 | 0 | 0 |
| hid | ɪ | 0 | 0 | 0 | 0 | 0 | 0 |
| head | ɛ | 0 | 1 | 0 | 0 | 0 | 0 |
| had | æ | 0 | 2 | 0 | 0 | 1 | 0 |
| hod | ɑ | 0 | 2 | 2 | 1 | 0 | 0 |
| hawed | ɔ | 0 | 1 | 2 | 0 | 1 | 0 |
| hood | u | 0 | 1 | 2 | 0 | 0 | 0 |
| who'd | u | 0 | 0 | 2 | 1 | 1 | 0 |
| hud | ʌ | 0 | 2 | 1 | 0 | 0 | 0 |
| heerd | ɪə | 1 | 0 | 0 | 1 | 1 | 0 |
| heard | ɜ | 1 | 1 | 1 | 0 | 1 | 0 |
| hard | ɑɜ | 1 | 2 | 2 | 1 | 1 | 0 |
| hoard | oɜ | 1 | 1 | 2 | 1 | 1 | 0 |
| haird | ɛɜ | 1 | 1 | 0 | 0 | 1 | 0 |
| heyed | eɪ | 0 | 1 | 0 | 1 | 1 | 1 |
| hoed | oʊ | 0 | 1 | 2 | 1 | 1 | 2 |
| hide | aɪ | 0 | 2 | 1 | 1 | 1 | 1 |
| how'd | aʊ | 0 | 2 | 1 | 1 | 1 | 2 |
| hoy'd | ɔɪ | 0 | 1 | 2 | 1 | 1 | 1 |

sensitivity setting: 3.45; range of subjects' settings: 2.5–4; total range: 1–8). Each subject used the processor at the same setting for the six sessions.

2. Evaluation with speech and warble tones

In a session prior to data collection, subjects responded to practice lists for vowels, consonants, words, and sentences at simulated soft, conversational, and raised-to-loud vocal efforts (50, 60, and 70 dB sound-pressure level). Test sessions were limited to two hours to minimize subject fatigue and inattention. To meet this criterion, order of testing was as follows: sessions 1 and 4—vowels, consonants, and CNC Words; sessions 2 and 5—vowels, consonants, and CUNY Sentences; sessions 3 and 6—CNC Words and CUNY Sentences. Within each test type, the three levels were pseudorandomized among the four test sessions. One list of vowels, consonants, and words and two lists of sentences were presented at each level during each session. To minimize variance associated with different word and sentence lists as well as fatigue effects, all subjects followed the same pseudorandom order of lists and presentation levels. That is, the experiment was replicated across subjects. In the second half of the study, the same randomizations of vowels and consonants and the CNC Word lists were presented as in the first half of the study but different pseudorandomizations of level were followed. Four weeks elapsed before a list or randomization was repeated. None of the sentence lists was presented a second time. Warble-tone thresholds were obtained only during the first and last sessions because variance associated with them is small (Skinner *et al.*, 1995b).

3. Analysis of data

Analysis of variance was used to determine whether performance at the three sound-pressure levels (and for each level paired with another) for the group and individual subjects was significantly different on each of the speech tests. In addition, data were analyzed to determine whether perfor-

mance during the first versus the second half of the study was significantly different. In these analyses, all effects, including subjects, were considered fixed, and the highest-order interaction was assumed to be zero.

Confusion matrices were created from group consonant and vowel data at each of the three levels (50, 60, and 70 dB sound-pressure level). For information transmission analysis (Miller and Nicely, 1955), vowels and consonants were coded according to the features as listed in Tables II and III. Based on acoustic analysis of the vowels, vowel height (first formant, *F*1) was categorized as 0, if *F*1 < 400 Hz; as 1, if *F*1 was between 400 and 600 Hz; and as 2, if *F*1 > 600 Hz. Vowel frontness (second formant, *F*2) was categorized as 0, if *F*2 > 1700 Hz; as 1, if *F*2 was between 1100 and 1700 Hz; as 2, if *F*2 < 1100 Hz. Vowel duration was categorized as 1, if it was greater than 250 ms, or 2, if it was less than 250 ms. Diphthongs were categorized as 2, if *F*2 dropped, and 1, if *F*2 rose. All other vowels were categorized as 0 for this feature. This ignores the fact that *F*2 also drops in, for example, “heerd” (/hiəɹd/), but it was considered that the major acoustic event in that word was the drop in *F*3, with the *F*2 drop a necessary consequence. Percent information transmitted was calculated for each level, both overall and for each feature of the vowels and consonants. Collapsed confusion matrices were formed for each feature, and a loglinear model and the likelihood-ratio chi-square statistic (G^2) were used to determine whether information transmitted was significantly different (Bell *et al.*, 1986) across all levels and pairs of the three levels.

The group mean and standard deviation of the warble-tone thresholds at each frequency as well as the standard error of measurement across subjects and frequencies were calculated. The standard error of measurement is the square root of the mean variance. The mean and range of thresholds across subjects for each frequency (Hz) were plotted on the same graph as the spectra of each speech material at the three overall levels for comparison.

TABLE III. Features specified for information transmission analysis.

| Consonants | | Features | | | | | | |
|------------|--------|----------|-------|-------------|-----------|-------|-----------|---------|
| Token | Symbol | Voiced | Nasal | Continuance | Frication | Place | Sonorance | Lateral |
| apa | p | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| aba | b | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| ata | t | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| ada | d | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| aka | k | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| aga | g | 1 | 0 | 0 | 0 | 3 | 0 | 0 |
| asa | s | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| aza | z | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| asha | ʃ | 0 | 0 | 1 | 1 | 2 | 0 | 0 |
| azha | ʒ | 1 | 0 | 1 | 1 | 2 | 0 | 0 |
| acha | tʃ | 0 | 0 | 0 | 1 | 2 | 0 | 0 |
| aja | dʒ | 1 | 0 | 0 | 1 | 2 | 0 | 0 |
| afa | f | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| ava | v | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| ama | m | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| ana | n | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| ala | l | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| ara | r | 1 | 0 | 1 | 0 | 1 | 1 | 0 |

II. RESULTS

A. MAP parameters

Subjects' MAP parameters, except for thresholds and maximum acceptable loudness levels (that is, minimum and maximum stimulation levels), are summarized in Table IV. Base level of 4 is the lowest setting; a base level of 6 was used for two subjects who find noise in everyday life bothersome. Seven of the ten subjects use the default Q value of 20; the other three find speech clearer with Q values of 25 or 30. Electrodes have been eliminated from subjects' MAPs for facial nerve stimulation (subjects 1, 3, and 4), aberrant or unpleasant sound (subjects 2, 6, and 7), or an open-circuit electrode (subject 7, electrode 9).

With the Spectra 22 speech processor and SPEAK speech coding strategy, frequency boundaries are assigned to electrodes by choosing from among a number of frequency boundary assignment tables (Technical Reference Manual,

1994). Although frequency boundary assignment Table 9 is the default setting in the fitting software, most subjects in this study used Table 7 with which frequencies from 120 to 1080 Hz are assigned to the six most apical electrodes. With Table 8, 133–1022 Hz are assigned to the five most apical electrodes. With Table 9, 150–950 Hz are assigned to the four most apical electrodes. Table 7 allows more electrodes, with presumably different pitch percepts, to be assigned to the first formant region for vowels and sonorant consonants than Tables 8 or 9 do. Tables 8 or 9 were used only when subjects found sound was too high-pitched with Table 7. For all tables, frequency boundaries for each band are preset; if fewer than 20 electrodes are used, the highest frequency causing stimulation is lower. For example, the upper frequency for Table 7 is 8656 Hz with 20 electrodes and 4595 Hz with 16 electrodes. To create a MAP that had acceptable balance between low- and high-pitched sound for subject 1,

TABLE IV. Subjects' MAP settings for electrode pairing mode, base level, Q value, number of active electrodes, frequency boundary assignment table, and preamplifier gain/electrodes.

| Subject | Electrode pairing mode | Base level | Q value | Number active electrodes | Frequency assignment table | Preamplifier gain (G)/electrode (E) |
|---------|------------------------|------------|-----------|--------------------------|----------------------------|-------------------------------------|
| 1 | BP+3 | 4 | 25 | 15 | 9 | G8/E18-17, 14-5; G6/E4-2 |
| 2 | BP+1 | 4 | 25 | 16 | 7 | G8/E20-4 |
| 3 | BP+3 | 4 | 20 | 17 | 7 | G8/E18-17, 15-1 |
| 4 | BP+1 | 4 | 20 | 18 | 8 | G7/E20-19; G8/E18-15, 13, 11-1 |
| 5 | CG | 6 | 20 | 20 | 7 | G6/E22-19, 7-4; G8/E18-8 |
| 6 | BP+1 | 4 | 20 | 18 | 7 | G7/E20; G8/E19-3 |
| 7 | BP+1 | 4 | 30 | 16 | | G8/E20-10, 6-2 |
| 8 | BP+2 | 4 | 20 | 19 | 7 | G8/E19-1 |
| 9 | BP+1 | 4 | 20 | 19 | 7 | G8/E20-7; G7/E6; G6/E5-2 |
| 10 | BP+1 | 6 | 20 | 20 | 8 | G6/E20-17, 4-1; G8/E16-5 |

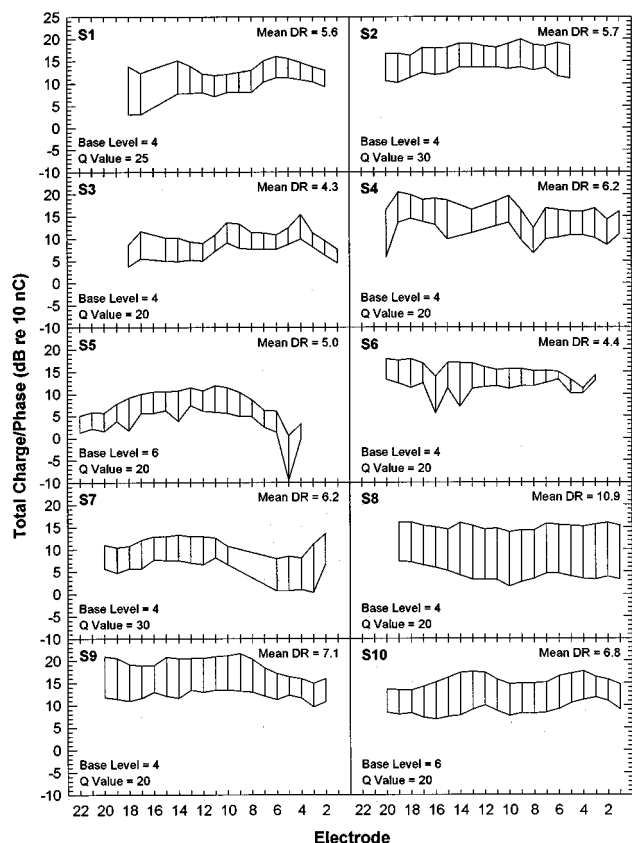


FIG. 1. MAP thresholds and maximum acceptable loudness levels (dB re: 10 nC) as the lower and upper contours, respectively, for each subject; mean dynamic range (DR; dB).

Table 9 was used and the bandwidth widened for electrodes 5 through 2 when the MAP was created. This widening was accomplished by double entry of each of these basal electrodes in the MAP so that the upper frequency was 9238 Hz. With single electrode entry, the upper frequency would have been 4903 Hz. Because high-frequency sound usually occurs at low-intensity levels in everyday life, these four electrodes were selected for stimulation less with the broader bandwidth (up to 9238 Hz). When these basal electrodes were stimulated less, the total percept of broadband sound was less high-pitched.

The default preamplifier gain in the fitting software is 8 on all electrodes. Gain is lowered on apical electrodes to decrease a booming quality and increase speech clarity; it is reduced on basal electrodes to decrease irritating high-pitched sound. Subjects in this study used a gain of 8 on most electrodes; two used it on all electrodes.

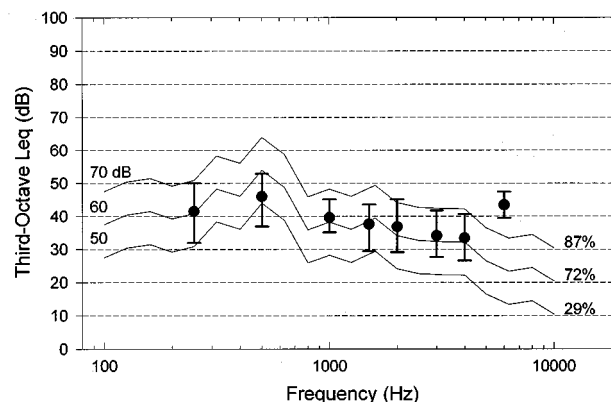


FIG. 2. Spectra of the CUNY Sentences for overall levels of 70, 60, and 50 dB sound-pressure level with associated group mean scores. Filled dots are group mean warble-tone thresholds; error bars represent the range.

MAP thresholds (lower contour) and maximum acceptable loudness levels (upper contour) for each subject are shown in Fig. 1; these contours represent the minimum and maximum stimulation levels, respectively. Subject 1 uses Current Level amplitude coding; other subjects use Stimulus level amplitude coding with which loudness is controlled mainly by change in pulse width (see Skinner *et al.*, 1991 for a description). In this figure, Current and Stimulus Level values have been converted to total charge/phase (dB re: 10 nC) using the calibration table for each subject's internal receiver/stimulator. There are marked differences in the configuration of threshold and maximum acceptable loudness level contours of subjects' MAPs.

B. Warble-tone thresholds

The group mean and standard deviation of the warble-tone thresholds, and the minimum and maximum thresholds are shown in Table V. Thresholds at 6000 Hz were not included in the mean for subjects 2, 3, and 7 because the upper frequency of incoming sound to their most basal electrode was lower than 6000 Hz. Standard error of measurement across subjects and frequencies was 2.2 dB which is close to the 2-dB stepsize of the threshold determination procedure.

C. Analysis of CUNY sentences recognition

The speech spectra (third-octave band L_{eq} , dB) for CUNY Sentences presented at 70, 60, and 50 dB sound-pressure level are shown in Fig. 2. At 70 dB, the overall L_{eq} was 67.4 dB (a long-term measure across all frequencies over four lists of 12 sentences). This L_{eq} value was lower

TABLE V. Group mean and standard deviation of the warble-tone thresholds (L_{eq} , dB) as well as minimum and maximum thresholds as a function of frequency (Hz).

| | Frequency (Hz) | | | | | | | |
|--------------------|----------------|------|------|------|------|------|------|------|
| | 250 | 500 | 1000 | 1500 | 2000 | 3000 | 4000 | 6000 |
| Mean | 41.5 | 46.0 | 39.6 | 37.6 | 36.8 | 34.1 | 33.4 | 43.4 |
| Standard deviation | 5.2 | 4.4 | 2.7 | 3.2 | 4.4 | 3.5 | 3.1 | 1.9 |
| Minimum | 32.0 | 36.9 | 35.1 | 29.5 | 29.1 | 27.7 | 26.6 | 39.4 |
| Maximum | 50.0 | 52.9 | 45.1 | 43.5 | 45.1 | 41.7 | 40.6 | 47.4 |

because 70 dB sound-pressure level was calculated using the average across the major peaks of the sentences; the majority of levels within a sentence fell below this peak.

The group mean scores are shown to the right of each contour; they are 87%, 72%, and 29% at 70, 60, and 50 dB sound-pressure level, respectively. There was a relatively small decrease in score (15 percentage points) from 70 to 60 dB, and a large decrease (43 percentage points) from 60 to 50 dB sound-pressure level. One standard error of the mean for these scores is 2.23%, 3.93%, and 4.30% at 70, 60, and 50 dB sound-pressure level, respectively. Mean score differences across the three levels and for pairs of levels (70/60, 60/50, 70/50 dB sound-pressure level) were highly significant [$F(2,9)=357.87$, $p<0.001$; $F(1,9)=42.06$, $p<0.001$; $F(1,9)=286.66$, $p<0.001$; $F(1,9)=933.98$, $p<0.001$, respectively]. There also were significant subject by level interactions ($p<0.001$, $p=0.005$, $p=0.024$, and $p<0.001$, respectively). Although there was a significant improvement in score from the first to the second half of the study for the group across levels and for the 60/50- and 70/50-dB pairs [$F(2,9)=5.17$, $p=0.003$; $F(1,9)=5.79$, $p=0.003$; $F(1,9)=5.75$, $p=0.004$], there was no significant improvement in score for individual subjects.

Group mean and range of warble-tone thresholds have been plotted on the same graph as the CUNY Sentence spectra. Except at 500 and 6000 Hz, mean thresholds intertwine with the contour at 60 dB sound-pressure level. From this graph, it would appear that sentences at 50 dB should be inaudible except at 500 Hz. To understand why subjects can recognize speech cues with a speech contour below warble-tone thresholds, several concepts must be considered. First, there is a 30-dB range of momentary fluctuations in speech energy, 12 dB above and 18 dB below each contour (French and Steinberg, 1947). Consequently, there are speech cues at levels well below each contour. Second, the warble tones and sentences were presented with a linear amplification system and measured on the L_{eq} setting of the sound-level meter with the microphone at the same reference point in the sound field. For this reason, their levels can be compared directly. Although their long-term, rms levels can be compared directly, the range of momentary fluctuations is much smaller for warble tones than sentences. Third, the Spectra 22 speech processor maps a maximum of 30 dB of the incoming sound onto the output range between the minimum (threshold) and maximum (MAL) stimulation level on each of a subject's active electrodes. The average electrical output range for the subjects in this study was between 4.3 and 10.9 dB (see Fig. 1). Consequently, all acoustic inputs were compressed at the output, including speech. Because the preamplifier gain (sensitivity control setting) was chosen by each subject to make conversational speech (approximately 60 dB sound-pressure level) comfortably loud, only a portion of the 30-dB range of momentary fluctuations for the 50 dB sound-pressure level input was within the 30-dB range processed by the Spectra 22. Nevertheless, some acoustic cues must have been above the minimum stimulation levels because some words were identified correctly by all except subjects 7 and 5. Fourth, speech energy at 6000 Hz is represented by a long-term average instead of the momentary peaks that are salient for

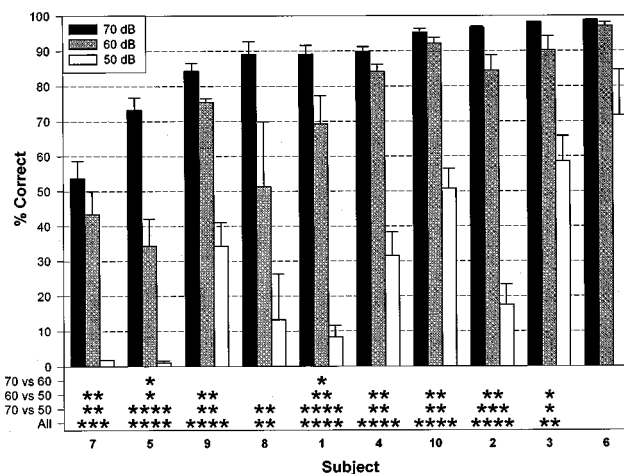


FIG. 3. Individual subjects' mean scores on the CUNY Sentences at 70, 60, and 50 dB sound-pressure level; error bars represent one standard error of the mean. Asterisks denote significantly different scores for pairs of sound levels or across all levels (****: 0.0001; ***: 0.001; **: 0.01; *: 0.05).

identifying /s/ and other high-frequency consonants. Consequently, it cannot be assumed that these high-frequency cues are not audible or recognizable. In fact, the group mean score for /s/ in the consonant test was 60% correct for an overall level of 50 dB sound-pressure level.

Individual subjects' mean scores on the CUNY Sentences across eight lists for each of the three levels are shown in Fig. 3. Analysis of variance was performed to determine if there were significant differences in scores for sound-level pairs and across sound levels. Asterisks at the bottom of Fig. 3 denote the significance level for conditions with significantly different scores. Eight out of ten subjects had scores that were not significantly different for the 70/60-dB pair, whereas all except subject 6, had significantly higher scores at 60 than at 50 dB sound-pressure level. At 70 dB sound-pressure level, scores ranged from 54% to 99% correct; at 60 dB, from 34% to 97%; and at 50 dB, from 1% to 72%. The marked decrease in score at 50 dB sound-pressure level was associated with many acoustic cues being soft or inaudible. However, subjects 5 and 7, who had chance scores, did not have higher warble-tone thresholds than subject 10, whose score was 51%. Subject 10 was able to recognize acoustic cues more easily near threshold than subjects 5 and 7. This performance difference probably occurred because subject 10 had greater ability to recognize speech at 60 and 70 dB sound-pressure level as well as more rapid growth of loudness near threshold than the other two subjects (Skinner *et al.*, unpublished data). Differences in the pattern of scores as a function of level for individual subjects (e.g., subjects 6 and 8) may also be related to their different growth of loudness patterns as well as where the minimum and maximum stimulation levels on each electrode were set in relation to the growth of loudness (see pilot data in Skinner *et al.*, in press).

D. Analysis of CNC word recognition

The speech spectra (third-octave band L_{eq} , dB) for CNC Words presented at 70, 60, and 50 dB sound-pressure level,

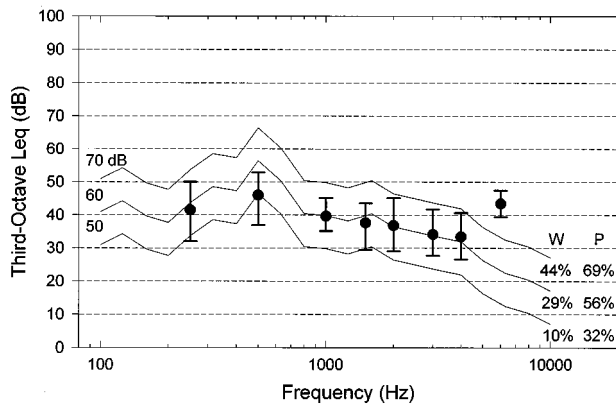


FIG. 4. Spectra of the CNC Words for overall levels of 70, 60, and 50 dB sound-pressure level with associated group mean scores. Filled dots are group mean warble-tone thresholds; error bars represent the range.

the mean word and phoneme scores, as well as the group mean and range of warble-tone thresholds are shown in Fig. 4. The configuration of these spectra is close to that for CUNY Sentences (Fig. 3); the L_{eq} of 69.3 dB is 2 dB higher. Given the calibration procedure, the majority of levels below the peak for each word is higher than for sentences. Group mean warble-tone thresholds intertwine with the 60-dB spectrum except at 500 and 6000 Hz.

The group mean word and phoneme scores are shown to the right of each contour in Fig. 4. The mean word and phoneme scores decrease less from 70 to 60 dB than from 60 to 50 dB sound-pressure level (14 and 13 percentage points for 70/60 dB; 19 and 24 percentage points for 60/50 dB). For 70, 60, and 50 dB sound-pressure level, one standard error of the mean for the words was 3.02%, 2.61%, and 1.52%, and

TABLE VI. Analysis of variance F ratios and levels of significance for CNC words and phonemes across 70, 60, and 50 dB sound-pressure level and for pairs of levels.

| | Words | | Phonemes | |
|------------|-----------|--------------|-----------|--------------|
| | F ratio | Significance | F ratio | Significance |
| All levels | 341.33 | <0.001 | 408.49 | <0.001 |
| 70/60 dB | 144.49 | <0.001 | 157.74 | <0.001 |
| 60/50 dB | 204.67 | <0.001 | 385.65 | <0.001 |
| 70/50 dB | 641.09 | <0.001 | 520.53 | <0.001 |

one standard error of the mean for the phonemes was 2.38%, 2.52%, and 2.71%, respectively. Mean word and phoneme score differences across the three levels and for pairs of levels (70/60, 60/50, 70/50 dB sound-pressure level) were highly significant as shown in Table VI. There were significant subject by level interactions across levels and for the 70/50 and 60/50 dB sound-pressure level pairs (words: $p < 0.001$, $p < 0.001$, $p = 0.001$; phonemes: $p = 0.005$, $p = 0.019$, $p = 0.017$, respectively). Word and phoneme scores for the first half of the study were not significantly different for the group across levels or for pairs of levels.

Individual subjects' mean CNC word and phoneme scores across four lists for each of the three levels are shown in Figs. 5 and 6. Asterisks at the bottom of these figures denote significance levels for sound-level pairs and across all levels (****: 0.0001; ***: 0.001; **: 0.01; *: 0.05).

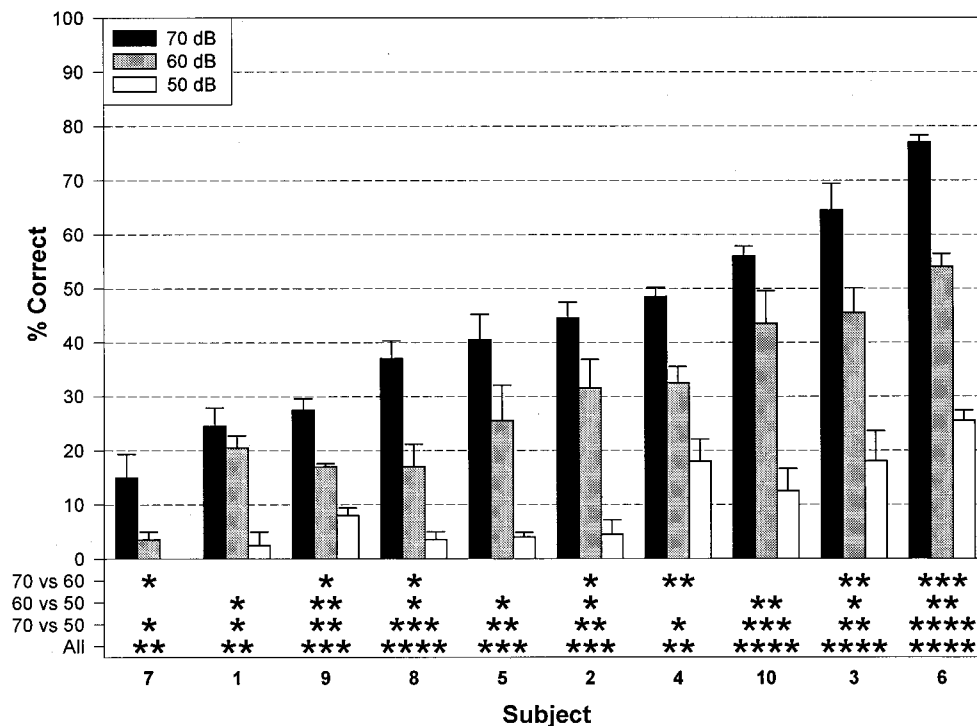


FIG. 5. Individual subjects' mean word scores on the CNC Words at 70, 60, and 50 dB sound-pressure level; error bars represent one standard error of the mean. Asterisks denote significantly different scores for pairs of sound levels or across all levels (****: 0.0001; ***: 0.001; **: 0.01; *: 0.05).

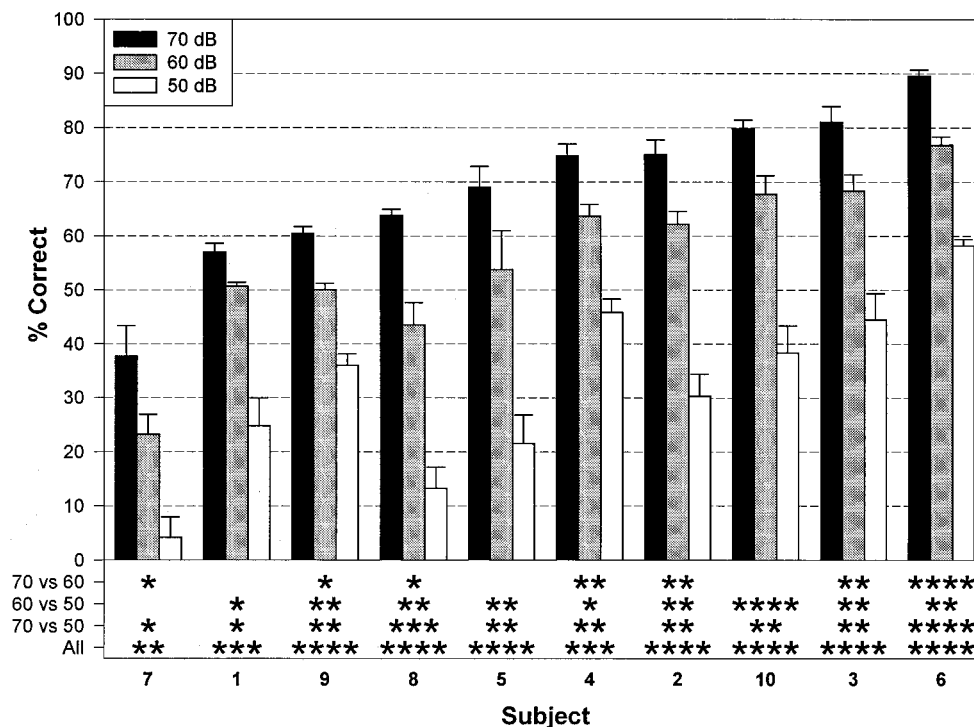


FIG. 6. Individual subjects' mean phoneme scores on the CNC Words at 70, 60, and 50 dB sound-pressure level with significance levels for each pair of levels and across levels. Error bars represent one standard error of the mean.

decrease in score at 50 dB sound-pressure level was associated with many acoustic cues being soft or inaudible.

E. Analysis of vowel recognition

The speech spectra (third-octave band L_{eq} , dB) for vowels presented at 70, 60, and 50 dB sound-pressure level, the mean scores, as well as the group mean and range of warble-tone thresholds are shown in Fig. 7. At each level, the vowel spectrum intersects that for the CUNY Sentences (see Fig. 2) only at 250 Hz; for frequencies below this, the vowel spectrum is as much as 10 dB lower, and for frequencies above this it is as much as 10 dB higher. The overall L_{eq} of the vowels is 72.9 dB, 5.5 dB higher than for the sentences and 3.6 dB higher than for the words. The reason for these dif-

ferences in overall L_{eq} values is that the levels below the major peak for vowels occur closer to the peak than for sentences and words. The group mean warble-tone thresholds intertwine with the 50-dB contour except at 250, 4000, and 6000 Hz. Consequently, more of the vowel acoustic cues were audible and recognizable than for sentences or words.

The group mean scores at 70, 60, and 50 dB sound-pressure level are shown at the right of the contours in Fig. 7; they are 64%, 70%, and 56%, respectively. Unlike sentence and word scores, the vowel score at 60 dB is six percentage points higher than that at 70 dB. For 70, 60, and 50 dB sound-pressure level, one standard error of the mean for the vowels was 3.09%, 3.09%, and 2.95%, respectively. There were highly significant differences in group mean score across the three levels and for the pairs of 70/60, 60/50, and 70/50 dB sound-pressure level [$F(2,9)=59.50$, $p<0.001$; $F(1,9)=21.91$, $p<0.001$; $F(1,9)=92.30$, $p<0.001$; $F(1,9)=47.11$, $p<0.001$, respectively]. There was a significant improvement in group mean score between the first and second half of the study across levels and for the 70/60- and 60/50-dB pairs [$F(2,9)=4.84$, $p=0.005$; $F(1,9)=4.88$, $p=0.008$; $F(1,9)=3.90$, $p=0.019$, respectively]. Across levels and for the 70/50- and 60/50-dB pairs, there were significant subject by level interactions ($p<0.001$). Across levels and for the 70/50-dB pair, there were significant subject by improvement in score interactions ($p=0.004$ and 0.037 , respectively). Because the six randomizations all included the same tokens, it is assumed that learning occurred during the study. There was a significant improvement in group sentence scores in the second half of the study; however, no sentence lists were repeated as they were for the vowels. There was no significant improvement in CNC Word or, as

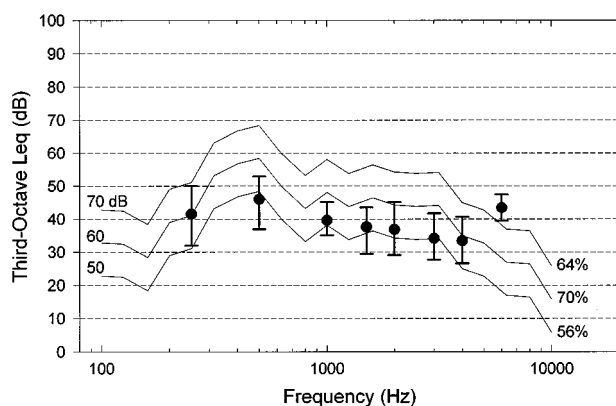


FIG. 7. Spectra of the vowels for overall levels of 70, 60, and 50 dB sound-pressure level with associated group mean scores. Filled dots are group mean warble-tone thresholds; error bars represent the total range.

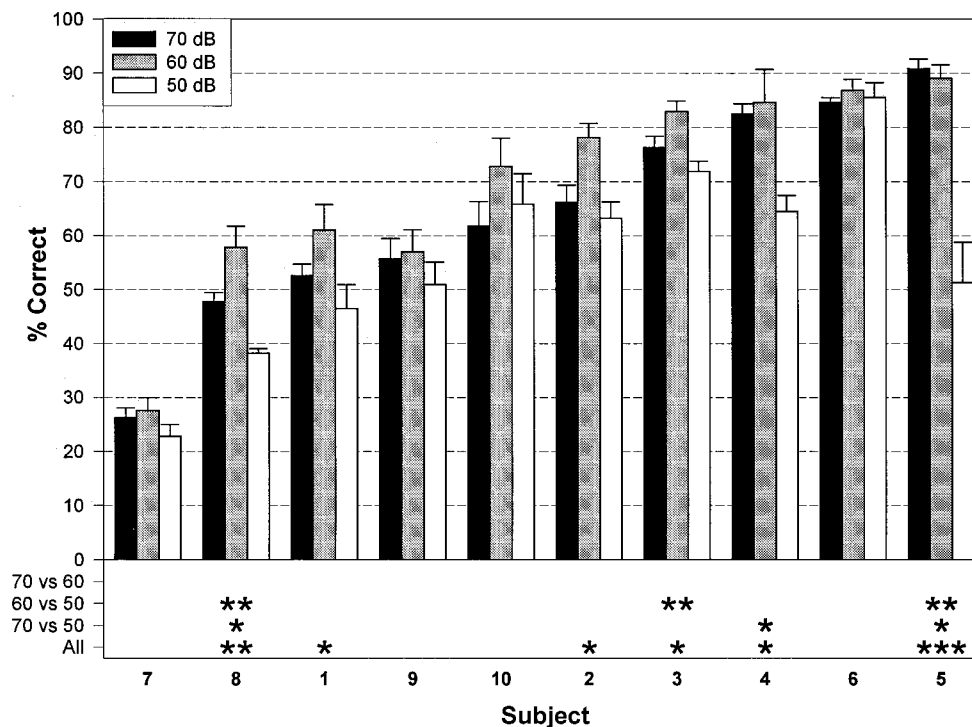


FIG. 8. Individual subjects' mean vowel scores at 50, 60, and 70 dB sound-pressure level; error bars represent one standard error of the mean. Asterisks denote significantly different scores for pairs of sound levels or across sound levels (***: 0.001; **: 0.01; *: 0.05).

described in the next section, in consonant scores.

Individual subjects' mean vowel scores across four lists for each of the three levels are shown in Fig. 8. Asterisks at the bottom of these figures denote significance levels for sound-level pairs and across sound levels for conditions with significantly different scores. At 70 dB sound-pressure level, vowel scores range from 26% to 91%; at 60 dB, from 28% to 89%; and at 50 dB, from 23% to 86%. Although the group mean score at 60 dB was significantly higher than that at 70 dB sound-pressure level, no individual subject had a significantly different score. Only three subjects had a significantly lower score at 50 than at 60 or 70 dB sound-pressure level. The relatively high scores at 50 dB, compared to words and sentences, were due to audibility and greater loudness of more acoustic cues.

Although individual subjects did not have a significantly higher score at 60 than 70 dB sound-pressure level, it is important to explore for reasons why the group mean at 60 dB was significantly higher than at 70 dB sound-pressure level. It is possible that vowel energy at 70 dB was limited somewhere between the sound-field amplification system and stimulation of subject's electrodes. The linear relation between third-octave vowel spectra measured at 60, 70, and 80 dB sound-pressure level and lack of audible distortion suggest that the vowels were not limited in the amplification system. Although subjects' optimum sensitivity control settings on the speech processor were set so that running speech at 70-dB A-weighted sound level was not supposed to activate the automatic gain control, some acoustic energy in the vowels at the 70-dB presentation level probably was compressed. In addition, vowels were compressed with the mapping function of incoming sound onto each subject's electri-

cal dynamic range between minimum and maximum stimulation levels (Q value). With Q values of 20, 25, and 30, the upper 10 dB of the 30-dB range of incoming sound is mapped onto the upper 80, 75, or 70% of this dynamic range for each electrode.

CIDIAG software developed by Bögli *et al.* (1994) and Dillier *et al.* (1994) was used to evaluate the effect of the automatic gain control that followed the preamplifier in the circuit and the nonlinear, Q -value mapping function on the electrical stimulation pattern in response to the vowel in the word, "who'd." This vowel was chosen because it had the greatest decrement in score between 70 and 60 dB sound-pressure level (see Table VII). Acoustically, this vowel's first and second formants ($F1$ and $F2$) as well as the majority of its spectral energy were below 1000 Hz. Analysis was done for four subjects, two of whom (subjects 3 and 8) had "who'd" identification scores that were 75 percentage points lower at 70 than 60 dB sound-pressure level, whereas, the other two (subjects 4 and 6) had scores that were 100% at both levels. "Who'd" was presented at these two levels to the implant microphone of the Spectra 22 speech processor programmed with each of the four subject's MAPs and the sensitivity control setting used for testing. The resulting digital code for each subject and level was transmitted to the implant receiver/stimulator in the Dual Processor Interface (Cochlear Pty. Ltd.); this code was captured and analyzed with the CIDIAG software.

To evaluate the effect of level on electrode stimulation, a pair of histograms of the weighted number of stimulations that occurred on each of the electrodes in a subject's MAP in response to the vowel in "who'd" presented first at 60 dB and then at 70 dB sound-pressure level were created. In these

TABLE VII. Group mean score (percent correct) for each vowel in the vowel test presented at 50, 60, and 70 dB SPL ordered from lowest to highest score at 70 dB SPL.

| Token | Symbol | 50 dB | 60 dB | 70 dB |
|-------|--------|-------|-------|-------|
| hod | a | 26.7 | 34.2 | 26.7 |
| hawed | ɔ | 16.7 | 41.7 | 42.5 |
| who'd | u | 80.0 | 76.7 | 48.3 |
| hard | ɑɹ̥ | 23.3 | 49.2 | 51.7 |
| heyed | eɪ | 54.2 | 60.0 | 54.2 |
| hid | ɪ | 41.7 | 59.2 | 56.7 |
| heerd | ɪɹ̥ | 55.8 | 75.8 | 60.8 |
| hoed | oʊ | 55.8 | 60.8 | 61.7 |
| hoy'd | ɔɪ | 50.8 | 61.7 | 61.7 |
| heed | i | 69.2 | 75.8 | 64.2 |
| hide | aɪ | 51.7 | 71.7 | 64.2 |
| how'd | aʊ | 65.8 | 80.8 | 65.8 |
| haird | ɛɹ̥ | 67.5 | 83.3 | 68.3 |
| hoard | oɹ̥ | 61.7 | 76.7 | 71.7 |
| head | ɛ | 68.3 | 75.0 | 75.8 |
| hud | ʌ | 60.0 | 79.2 | 80.8 |
| had | æ | 71.7 | 80.8 | 82.5 |
| hood | ʊ | 67.5 | 87.5 | 93.3 |
| heard | ɜɹ̥ | 76.7 | 95.8 | 94.2 |

histograms, each pulse on an electrode was weighted by the percentage of the electrical dynamic range at which it was stimulated (expressed as a multiplier/factor between 0 and 1). This weighted value of each pulse was then cumulated to a total for each electrode; these totals provided a better representation of the audibility and loudness of the pulses than if unweighted values were used. These two histograms are shown in Fig. 9 for subject 8, whose scores at 60 and 70 dB sound-pressure level were 92% and 17%, respectively. There were 18 to 33 more weighted stimulations on electrodes 19–14 and 8 in the 70-dB histogram than on the same electrodes in the 60-dB histogram. There were no stimulations on electrodes 13, 12, 9, 7, 6, and 5 in the 60-dB histogram because the processed output would have been below subject 8’s minimum stimulation level on these electrodes. Total weighted stimulations on these electrodes in the 70-dB histogram were between 1 and 11. Pairs of histograms obtained for subjects 3, 4, and 6 showed the same patterns of electrode

stimulation at the two levels. These results suggest that stimulation was substantially greater on the electrodes to which the formant frequencies ($F1$, $F2$, and $F3$) were assigned than at adjacent electrodes for both 60 and 70 dB sound-pressure level input levels.

To evaluate the effect of mapping input sound-pressure level onto the output dynamic range of stimulation, histograms of the occurrence of stimulation across all electrodes as a function of percent dynamic range between minimum and maximum stimulation levels in response to “who’d” were plotted for 60 and 70 dB sound-pressure level. A pair of these histograms is shown for subject 8 in Fig. 10. At 60 dB, electrode stimulation rarely occurred at or above 90% of the dynamic range, whereas, at 70 dB, the occurrence of stimulation at 100% of the dynamic range surpassed that at any other percentage. Pairs of histograms obtained for the other three subjects showed the same patterns. These results suggest that “who’d,” presented at 70 dB, caused more stimulation near or at the top of subjects’ electrical dynamic range than at 60 dB sound-pressure level.

The similarity of electrical stimulation patterns across subjects coupled with the very different scores for the two pairs of subjects suggests that they differed in how they perceived the acoustic cues for “who’d” from the electrical stimulation patterns they received. That is, stimulation at or near the maximum level for the 70-dB level did not interfere with identification of “who’d” for subjects 4 and 6 but it did for subjects 3 and 8. As charge/phase on individual electrodes increased for “who’d” presented at 60 to 70 dB sound-pressure level (as shown by the increase in number of weighted stimulations in Fig. 9 and greater stimulation at higher percentages of the dynamic range in Fig. 10), perceived growth of loudness may have changed in a nonlinear manner and its effect on recognition of “who’d” may have differed substantially for the two pairs of subjects. At 60 dB, it is possible that stimulation on the electrode to which $F1$ was assigned dominated the loudness perception (see electrode 18 for subject 8 in Fig. 9); at 70 dB, there may have been an increase of loudness perception from stimulation of electrodes adjacent to the $F1$ and $F2$ electrodes that was

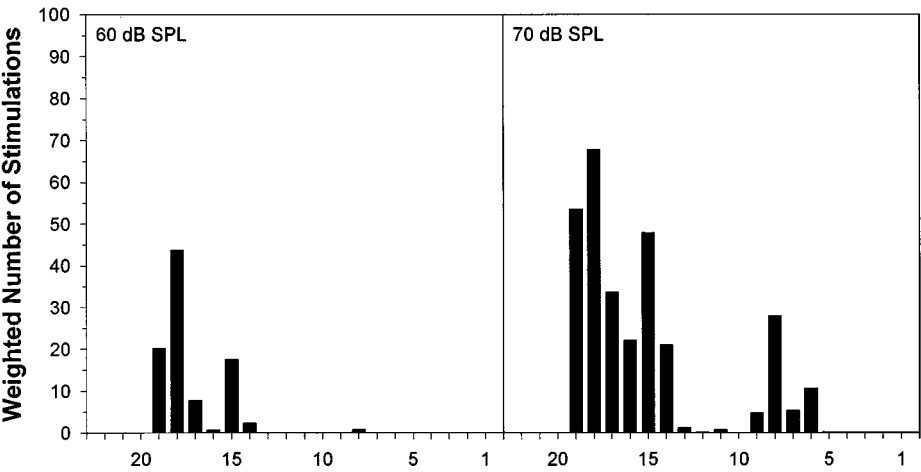


FIG. 9. Histograms of the weighted number of stimulations on each electrode obtained with the CIDIAG program for “who’d” presented at 60 and 70 dB sound-pressure level (SPL) and processed with the Spectra 22 speech processor programmed with subject 8’s MAP.

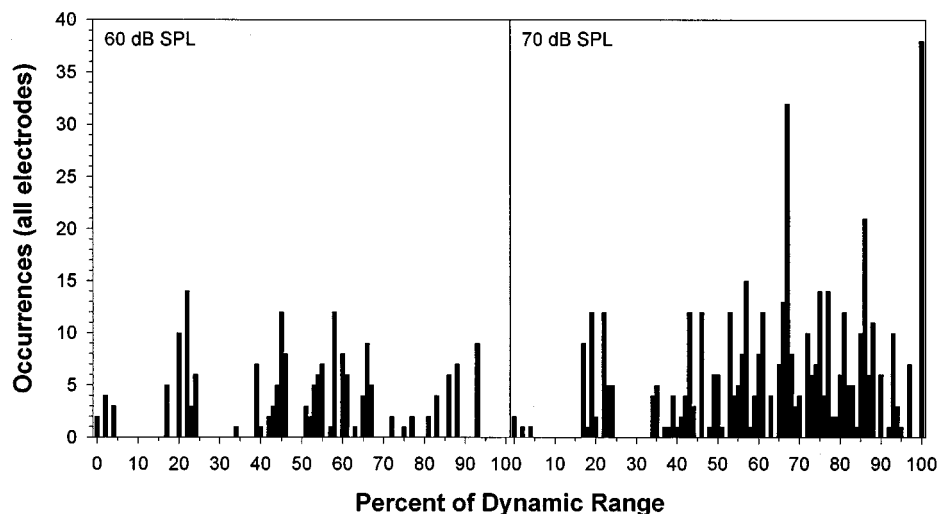


FIG. 10. Histograms of the occurrences of stimulation on all electrodes as a function of the percent dynamic range between minimum and maximum stimulation levels on each electrode obtained with the CIDIAG program for “who’d” presented at 60 and 70 dB sound-pressure level (SPL) and processed with the Spectra 22 speech processor programmed with subject 8’s MAP.

associated with either maintenance or decrement in recognition of “who’d.” Further research on this complex issue is needed.

The total information transmitted at 70, 60, and 50 dB sound-pressure level was 2.49, 2.72, and 2.21 bits, respectively. Percent information transmitted for the three levels is shown for the six vowel features in Fig. 11. Asterisks at the bottom of the figure denote significance levels for sound-level conditions that were significantly different. The significance levels were calculated using a loglinear model and the maximum-likelihood chi-square statistic (G^2). Percent infor-

mation transmitted was significantly higher at 60 than at 70 and 50 dB sound-pressure level except for the front ($F2$) and diphthong features for the 70/60-dB pair. Percent information transmitted was higher for the front ($F2$) feature than for the other features.

F. Analysis of consonant recognition

The speech spectra (third-octave band L_{eq} , dB) for consonants presented at 70, 60, and 50 dB sound-pressure level, the mean scores, as well as the group mean and range of

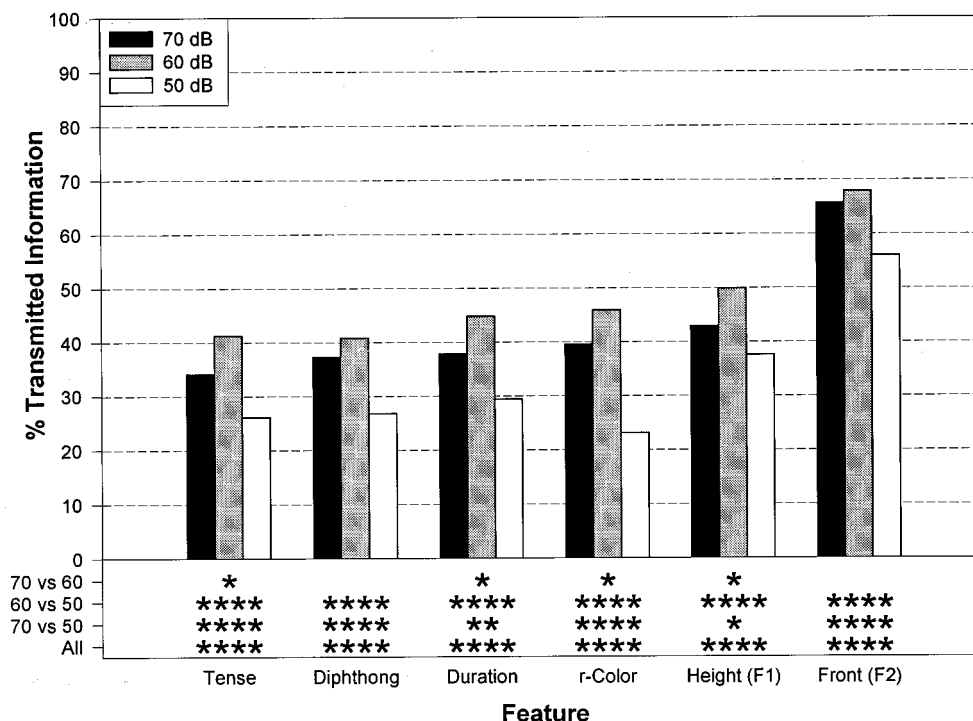


FIG. 11. Percent information transmitted for the six vowel features [r -color, height ($F1$), front ($F2$), tense, duration, and diphthong]. Asterisks denote significant differences in percent information transmitted for pairs of sound levels or across sound levels (****: 0.0001; **: 0.01; *: 0.05).

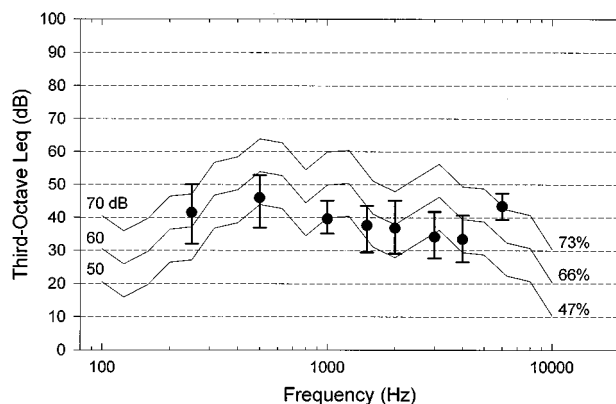


FIG. 12. Spectra of the consonants for overall levels of 70, 60, and 50 dB sound-pressure level with associated group mean scores. Filled dots are group mean warble-tone thresholds; error bars represent the range.

warble-tone thresholds are shown in Fig. 12. These spectra intersect those for the vowels (see Fig. 7) between 600 and 1000 Hz and at 3000 Hz; the consonant spectra are lower between 250 and 500 Hz, and between 1500 and 2000 Hz; they are slightly higher above 3000 Hz. The L_{eq} of the consonants was 69.9 dB, 2.5 dB higher than the sentences, 0.6 dB higher than the words, and 3.0 dB lower than the vowels. The reason for these differences in L_{eq} values is as follows. The levels below the major peak for the consonants (/aCa/) and words (CNC) occur about the same level. The vowels before and after the consonants cause the levels below the major peak to be higher than those in sentences in which all except the one or two stressed words are spoken at lower levels. Although there are two vowels surrounding the consonant in the consonant tokens (/aCa/) and one vowel for the

vowel tokens (/hVd/), the vowels in the vowel tokens are more stressed and of longer duration than the two vowels in the consonant tokens (/aCa/). In addition, stressed vowels are almost always more intense than stressed consonants. For these reasons, the vowel test L_{eq} was higher than the consonant test L_{eq} . The group mean warble-tone thresholds intertwine with the 50-dB contour except at 2 and 6 kHz. Although it appears that more of the consonant acoustic cues are audible than for sentences or words, the measured sound-pressure level is strongly dependent on the vowel energy.

The group mean scores at 70, 60, and 50 dB sound-pressure level are shown at the right of the contours in Fig. 12; they are 73%, 66%, and 47%, respectively. Like sentence and word scores, there is a smaller decrease in consonant scores from 70 to 60 dB (7 percentage points) than from 60 to 50 dB sound-pressure level (19 percentage points). For these three levels, one standard error of the mean for the consonants was 3.31%, 3.49%, and 3.40%, respectively. There were highly significant differences in group mean score across the three levels and for the pairs of 70/60, 60/50, and 70/50 dB sound-pressure level [$F(2,9)=290.41$, $p<0.001$; $F(1,9)=63.85$, $p<0.001$; $F(1,9)=314.04$, $p<0.001$; $F(1,9)=381.85$, $p<0.001$, respectively]. There was no significant improvement in group mean score between the first and second half of the study. Across levels and for the 70/60- and 70/50-dB pairs, there were significant subject by level interactions ($p<0.001$).

Individual subjects' mean consonant scores across four lists for each of the three levels are shown in Fig. 13. Asterisks at the bottom of the figure denote significance levels for sound-level pairs and across sound levels for conditions with significantly different scores. At 70 dB sound-pressure level,

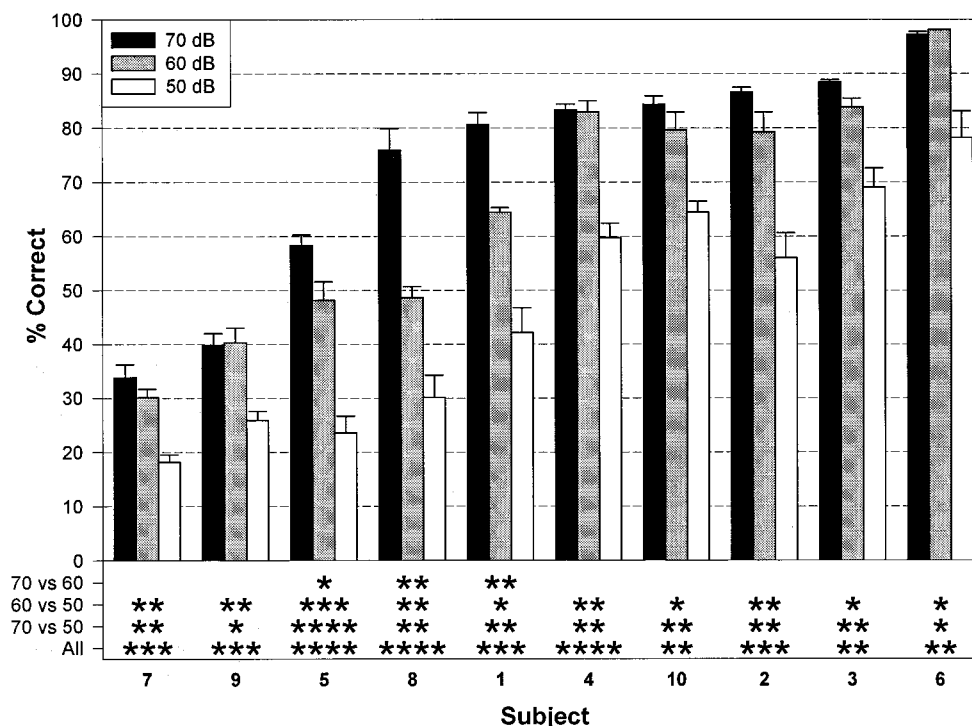


FIG. 13. Individual subjects' mean consonant scores at 50, 60, and 70 dB sound-pressure level; error bars represent one standard error of the mean. Asterisks denote significantly different scores for pairs of sound levels or across sound levels (****: 0.0001; ***: 0.001; **: 0.01; *: 0.05).

TABLE VIII. Group mean score (percent correct) for each consonant in the consonant test presented at 50, 60, and 70 dB SPL ordered from lowest to highest score at 70 dB SPL.

| Token | Symbol | 50 dB | 60 dB | 70 dB |
|-------|--------|-------|-------|-------|
| ava | v | 22.5 | 54.2 | 30.0 |
| ala | l | 26.7 | 45.0 | 52.5 |
| acha | tʃ | 50.8 | 55.0 | 56.7 |
| ada | d | 33.3 | 47.5 | 57.5 |
| azha | ʒ | 40.0 | 54.2 | 65.8 |
| ara | r | 47.5 | 71.7 | 72.5 |
| ama | m | 46.7 | 61.7 | 74.2 |
| aga | g | 50.8 | 70.0 | 76.7 |
| aza | z | 41.7 | 68.3 | 77.5 |
| aka | k | 65.8 | 72.5 | 78.3 |
| aba | b | 56.7 | 75.0 | 78.3 |
| apa | p | 27.5 | 63.3 | 81.7 |
| asha | ʃ | 41.7 | 75.8 | 82.5 |
| asa | s | 60.0 | 75.8 | 83.3 |
| aja | dʒ | 76.7 | 80.0 | 83.3 |
| ana | n | 50.0 | 74.2 | 85.0 |
| afa | f | 32.5 | 54.2 | 86.7 |
| ata | t | 70.0 | 80.8 | 88.3 |

consonant scores range from 34% to 97%; at 60 dB, from 30% to 98%; and at 50 dB, from 18% to 78%. As shown in Fig. 12, scores were significantly different across levels and were significantly higher at 70 than 50 dB and at 60 than 50 dB sound-pressure level for all subjects. Subjects 1, 5, and 8 had significantly higher scores at 70 than at 60 dB sound-pressure level.

The group mean scores for each consonant at the three levels are shown in Table VIII. These scores represent the diagonal entries of the three confusion matrices. The total

information transmitted for consonants at 70, 60, and 50 dB sound-pressure level was 2.90, 2.55, and 1.73 bits, respectively. Percent information transmitted for the three levels is shown for the seven consonant features in Fig. 14. Asterisks at the bottom of the figure denote significance levels for conditions with significantly different percent information transmitted values. The significance levels were calculated using a loglinear model and the maximum-likelihood chi-square statistic (G^2). Percent information transmitted was significantly higher across all levels and all pairs except for the laterals and for the 70/60 pair of the sonorance feature. Percent information transmitted was higher for the sonorance feature than for the other features at 70 and 60 dB sound-pressure level. It was lowest for the lateral feature.

III. DISCUSSION

The major finding of this study is that these ten postlinguistically deaf adults using the Spectra 22 and SPEAK speech coding strategy recognized significant amounts of unrelated sentences, single syllable words, words scored as phonemes correct, vowels, and consonants at a conversational level of 60 dB sound-pressure level (e.g., group mean scores: 72%, 29%, 56%, 70%, and 66%, respectively) in a sound-only condition. In addition, all except subject 7 recognized some words at 50 dB sound-pressure level; this level simulates either soft speech or speech spoken at a normal conversational effort but at distances of approximately 3 m. We have observed that this level of speech recognition performance in a sound-only condition is associated with these subjects' ability to carry on fluent communication when speech is spoken at a normal level (60 dB measured with a

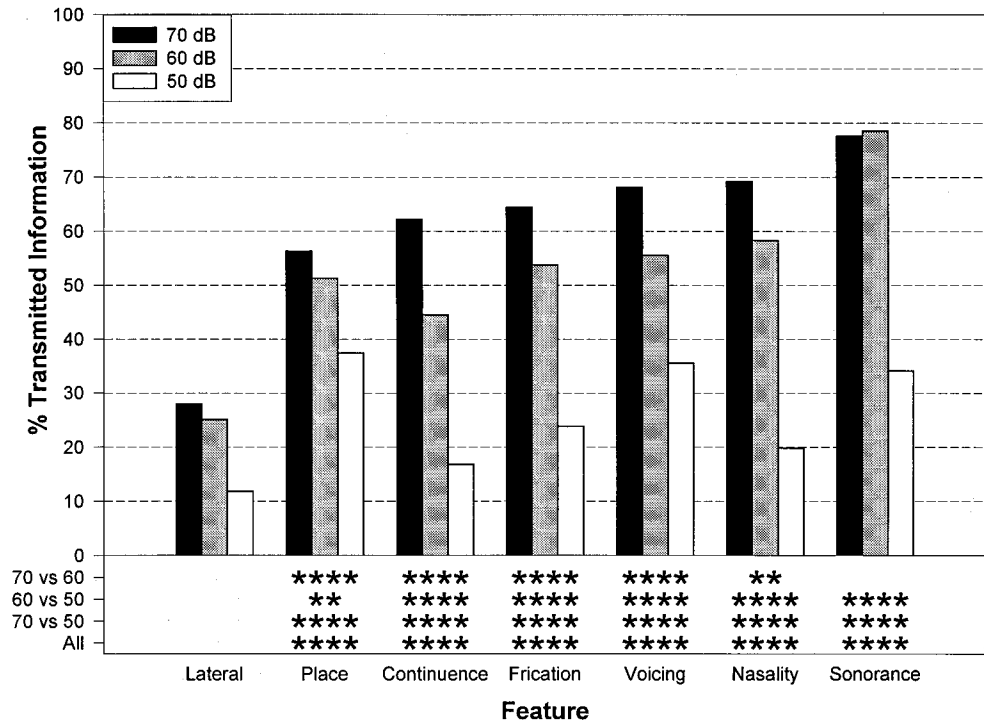


FIG. 14. Percent information transmitted for the seven consonant features (voicing, nasality, continuance, frication, place, sonorance, and lateral). Asterisks denote significant differences in percent information transmitted for pairs of sound levels or across sound levels (****: 0.0001; **: 0.01).

sound-level meter near the ear of the subjects) at approximately one meter with speechreading cues. Most of these subjects also understand much of what is said in similar situations without speechreading. Although they may understand only a portion of what is said at greater distances (and lower levels at the implant microphone), they hear enough to move closer or to ask the person to speak louder if they want to understand more. They can also increase or decrease the sensitivity control setting on the speech processor to listen to consistently softer or louder sounds. This access to speech and other sounds over a range of levels (50–70 dB sound-pressure level) and distances (over 3 m to a few inches when a telephone is used) enables them to communicate successfully with a variety of people in many situations. These practical benefits provided by the cochlear implant are essential to consider when evaluating potential candidates for implantation.

Measurement of speech spectra, overall levels, and output from several subjects' speech processors provided some information on why vowel scores were higher at 60 than 70 dB sound-pressure level, whereas, for the other materials, scores were higher at 70 dB. The original calibration of overall sound-pressure level was based on an average across peak levels within a list of vowels, consonants, words, and sentences. This method is similar to the use of frequent peaks on the VU meter to set the level of a calibration tone for a speech test. Subsequent L_{eq} measurement of overall levels revealed a 5.5-dB difference between that for sentences (67.4 dB) and vowels (72.9 dB) when all were presented at 70 dB sound-pressure level. Comparison of the L_{eq} long-term average level with the average across single peaks for a vowel or sentence list indicates that vowel levels cluster much closer to the peaks whereas the speech energy in a sentence often falls below that of the peak. For words, the long-term level (69.3 dB L_{eq}) is 2 dB closer to the average of the peaks than for the sentences. These calibration issues are important when analyzing speech recognition performance as a function of level for incoming sound, particularly because of the nonlinear mapping of this sound onto each subject's electrical dynamic range between minimum and maximum stimulation levels in the MAP.

This nonlinear mapping coupled with the growth of loudness in each subject's electrical auditory area probably accounts for the smaller decrease in score from 70 to 60 dB sound-pressure level and the larger decrease in score from 60 to 50 dB for consonants, words, and sentences. At 50 dB, only part of incoming speech energy was within the 30-dB range analyzed by the speech processor for the sensitivity control setting used for testing. This speech energy was mapped onto the lower part of the electrical dynamic range, and consequently, its loudness was perceived as very soft, soft, or medium soft depending on the individual subject's growth of loudness near threshold. In the study by Müller-Deile *et al.* (1995), the pattern of scores at 70, 60, and 50 dB sound-pressure level for German monosyllables and sentences was similar to that in the present study despite the difference in language, speech materials, and use of MPEAK instead of SPEAK. For example, mean Göttinger Sentence scores decreased six percentage points between 70 and 60 dB

sound-pressure level and 40 percentage points between 60 and 50 dB. In the present study, the mean CUNY Sentence scores decreased 15 and 43 percentage points for these two pairs of levels, respectively. This similarity in score differences for the two studies appears related to the use of the same nonlinear mapping function for both speech coding strategies.

The group mean sentence and word scores in this study (87% and 44%, respectively) are higher than those in the MPEAK/SPEAK field study (78% and 34%, respectively), for which the same recordings were used (Skinner *et al.*, 1994). There are several possible reasons for this difference. First, a wider range of abilities was probably represented by the 63 subjects in the field study than in the present study. Second, half of the subjects in the present study had 15 months experience with SPEAK prior to testing, whereas, those in the field study had a total of 5 weeks and 8 weeks prior to the two testing periods. Third, frequency boundary assignment Table 9 was used for all field-study subjects, whereas, only one of this study's subjects used Table 9 (7 subjects used frequency boundary assignment Table 7; 2 subjects used Table 8). Skinner *et al.* (1995a) have shown that a group of seven subjects' scores on vowels, consonants, words, and sentences were significantly higher with their processors programmed with frequency boundary assignment Table 7 than with Table 9.

Subjects who participated in the present study represent a range of speech recognition abilities that differ within one sound-pressure level and speech material as well as across levels or speech materials. For example, scores among the ten subjects range between 54% and 99% correct for sentences presented at 70 dB sound-pressure level. Another example is the difference in scores across levels for sentences (subjects 10 and 8). Subject 10's scores at 70 and 60 dB sound-pressure level were almost the same (95% and 92%) and lower at 50 dB (72%), whereas subject 8's scores were widely separated at the levels (89%, 51%, and 13%). It appears that subject 8 had more difficulty recognizing acoustic cues below the raised-to-loud level than subject 6. In a companion, unpublished study, subject 8 had slower growth of loudness functions on individual electrodes as well as across electrodes than subject 6. Pilot data showed that raising the minimum stimulation levels in subject 8's MAP provided her with better speech recognition at 60 dB sound-pressure level (33% improvement in score; Skinner *et al.*, in press). A third example is the different patterns of speech recognition across speech materials. Subject 5 had significantly higher vowel scores at 70 and 60 dB sound-pressure level (91% and 89%, respectively) compared to subject 1 (53% and 61%, respectively; two-tailed, t test: $p < 0.001$ and $p = 0.002$, respectively), whereas, she had significantly lower consonant scores at 70, 60, and 50 dB sound-pressure level (58%, 48%, and 24%, respectively) than subject 1 (81%, 64%, and 42%, respectively) (two-tailed, t test: $p = 0.0003$, $p = 0.0040$, and $p = 0.0156$, respectively). Subject 5 had a much wider dynamic range between threshold and maximum acceptable loudness level for apical than basal electrodes (see Fig. 1). Based on her performance, it appears that she recognized lower frequency acoustic cues, provided by electrical stimu-

lation of more apical electrodes, better than higher frequency cues, provided by more basal electrodes. In contrast, subject 1 had two apical electrodes (E16 and 15) eliminated from his MAP because they caused facial nerve stimulation secondary to otosclerosis. For him, first formant frequencies below 950 Hz were assigned to apical electrodes 18, 17, 14, and 13. Based on his performance, it appears that he obtained less low-frequency information for vowel recognition and more mid- and high-frequency information for consonant recognition. These two subjects had very similar word scores at 50 and 60 dB sound-pressure level (two-tailed, t test: $p=0.50$ and $p=0.59$, respectively), but subject 5's score was significantly higher at 70 dB (12% difference; two-tailed, t test: $p=0.0333$). Reasons for these individual variations need elucidation through further research on the relation between stimuli, sound processing, and subjects' responses. This research is needed to develop guidelines to select speech processing and MAP parameter values so that an individual's speech recognition with a cochlear implant can be optimized.

Several concepts should be considered in the selection of speech tests for evaluation of postlinguistically deaf adults as cochlear implant candidates. First, the speech test should reflect how well the patient understands speech by sound alone in *everyday life*. Open-set, sentence, and word tests approximate this situation more closely than closed-set vowel and consonant tests, even though the latter yield valuable information, particularly for research purposes. Second, the level at which speech tests are given clinically should represent that of conversational speech (i.e., approximately 60 dB sound-pressure level) and not the raised-to-loud speech level of 70 dB that is very difficult to maintain. The results of the present study suggest that many postlinguistically deaf adults understand a majority of the words in sentences and a number of individual words at this level. Third, ceiling and floor effects should be avoided. For the CNC Word test, scoring according to words correct avoids a ceiling effect for subjects in this study as well as patients who use the CIS strategy with the Clarion Cochlear Implant System (Kessler, personal communication) and the SPEAK strategy with the Nucleus Cochlear Implant System (Staller, personal communication). Scoring the same test according to phonemes correct avoids a floor effect (except for those patients who cannot recognize spectral cues or hear soft sounds). Fourth, variability should be minimized. Based on these four concepts and the results of the present study, it is suggested that a word and a sentence test presented at 60 dB sound-pressure level be considered for the preimplantation evaluation of postlinguistically deaf adults. Data from a much larger group of subjects would need to be obtained with these two tests at 60 dB sound-pressure level before appropriate guidelines (for scores in percent correct) can be established.

In summary, this limited group of ten adult implantees obtained high levels of open-set speech recognition at raised-to-loud and conversational levels (70 and 60 dB sound-pressure level, respectively), and all but one of them understood some words correctly at a soft level (50 dB) with the SPEAK speech coding strategy of the Nucleus 22 Channel Cochlear Implant System.

ACKNOWLEDGMENTS

Appreciation is expressed to the ten subjects who have contributed so much to this investigation. We are grateful to Norbert Dillier, Wai Kong Lai, Peter Seligman, Lesley Whitford, Winifred Strange, and two anonymous reviewers for their critical comments on an earlier draft of this paper. This research was supported by Grant No. R01 DC00581 from the National Institute on Deafness and Other Communication Disorders.

- Bell, T. S., Dirks, D. D., Levitt, H., and Dubno, J. R. (1986). "Loglinear modeling of consonant confusion," *J. Acoust. Soc. Am.* **79**, 518–525.
- Bogli, H., Dillier, N., Lai, W. K., and Rohner, M. (1994). CIDIAG for Windows: Diagnostic program for Cochlear Implant Speech Processor Signals (Version 8.0), Department of Otolaryngology, University Hospital, Zurich.
- Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (1988). "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," *Ear Hear.* **9**, 306–312.
- Byrne, D., and Dillon, H. (1986). "The National Acoustics Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid," *Ear Hear.* **7**, 257–265.
- Byrne, D., and Tonisson, W. (1976). "Selecting the gain of hearing aids for persons with sensorineural hearing impairments," *Scand. Audiol.* **5**, 51–59.
- Clark, G. M., Blamey, P. J., Brown, A. M., Busby, P. A., Dowell, R. C., Franz, B. K.-H., Pyman, B. C., Shepherd, R. K., Tong, Y. C., Webb, R. L., Hirshorn, M. S., Kuzma, J., Mecklenburg, D. J., Money, D. K., Patrick, J. F., and Seligman, P. M. (1987). "The University of Melbourne-Nucleus Multi-Electrode Cochlear Implant," *Adv. Otorhinolaryngol.* **38** (Suppl.), 1–189.
- Cox, R. M. (1983). "Using ULCL measures to find frequency/gain and SSPL90," *Hear. Instrum.* **34**(7), 17–21, 39.
- Dillier, N., Lai, W., and Bögli, H. (1994). "A high spectral transmission coding strategy for a multi-electrode cochlear implant," in *Advances in Cochlear Implants*, edited by I. J. Hochmair-Desoyer and E. S. Hochmair (International Interscience Seminars, Vienna), pp. 152–157.
- Dowell, R. C., Seligman, P. M., Blamey, P. J., and Clark, G. M. (1987). "Speech perception using a two-formant 22-electrode cochlear prosthesis in quiet and in noise," *Acta Otolaryngol.* **104**, 439–446.
- French, N. R., and Steinberg, J. C. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **19**, 90–119.
- Kessler, D. K. (1996). Personal communication.
- Miller, G. A., and Nicely, P. E. (1955). "An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.* **27**, 338–352.
- Müller-Deile, J., Schmidt, B. J., and Rudert, H. (1995). "Effects of noise on speech discrimination in cochlear implant patients," *Ann. Otol. Rhinol. Laryngol.* **104** (Suppl. 166), 303–306.
- Owens, E., Kessler, D. K., Raggio, M. W., and Schubert, E. D. (1985). "Analysis and revision of the Minimal Auditory Capabilities (MAC) Battery," *Ear Hear.* **6**, 280–290.
- Pascoe, D. P. (1978). "An approach to hearing aid selection," *Hear Instrum.* **29**(6), 12–16, 36.
- Pearsons, K. S., Bennett, R. L., and Fidell, S. (1976). "Speech levels in various environments," Bolt Beranek and Newman Report No. 321, Canoga Park, CA.
- Peterson, G., and Lehiste, I. (1962). "Revised CNC lists for auditory tests," *J. Speech Hear. Disord.* **27**, 62–70.
- Reid, L., Smiarowski, R. A., and McPherson, L. T. (1977). "Hearing aid gain-setting performance of experienced users," *Arch. Otolaryngol.* **103**, 203–205.
- Seligman, P., and McDermott, H. (1995). "Architecture of the Spectra 22 Speech Processor," *Ann. Otol. Rhinol. Laryngol.* **104** (Suppl. 166), 139–141.
- Seligman, P., and Whitford, L. (1995). "Adjustment of appropriate signal levels in the Spectra 22 and Mini Speech Processors," *Ann. Otol. Rhinol. Laryngol.* **104** (Suppl. 166), 172–175.
- Skinner, M. W. (1994). "The borderline child," Paper presented at the 5th Symposium on Cochlear Implants in Children, New York, NY (unpublished).

- 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. J., and Beiter, A. L. (1994). "Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System," *Am. J. Otol.* **15**, Suppl. 2, 15–27.
- Skinner, M. W., Holden, L. K., and Holden, T. A. (1995a). "Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy," *Ann. Otol. Rhinol. Laryngol.* **104** (Suppl. 166), 307–311.
- Skinner, M. W., Holden, L. K., and Holden, T. A. (in press). "Parameter selection to optimize speech recognition with the Nucleus implant," *Otolaryng. Head Neck Surg.*
- Skinner, M. W., Holden, L. K., and Holden, T. A. (1995). "Relation of speech recognition to mapping of sound onto the electrical auditory area of adults with cochlear implants," paper presented at the 1995 Conference on Implantable Auditory Prostheses, Pacific Grove, CA (unpublished abstract).
- Skinner, M. W., Holden, L. K., Holden, T. A., and Demorest, M. E. (1995b). "Test-retest measures of performance stability in adults with cochlear implants," *Ear Hear.* **16**, 187–197.
- 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 Mini Speech Processor (MSP) of the Nucleus Multi-Electrode Cochlear Implant," *Ear Hear.* **12**, 3–22.
- Staller, S. J. (1996). Personal communication.
- Technical Reference Manual* (1994). (Cochlear Corporation, Englewood, CO).
- Walden, B. E., Schuchman, G. I., and Sedge, R. K. (1977). "The reliability and validity of the comfort level method of setting hearing aid gain," *J. Speech Hear. Dis.* **42**, 455–461.
- Walker, G., Dillon, H., and Byrne, D. (1984). "Sound-field audiometry: recommended stimuli and procedures," *Ear Hear.* **5**, 13–21.

Propagation of damselfish (*Pomacentridae*) courtship sounds

David A. Mann^{a)}

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Phillip S. Lobel

Boston University Marine Program, Marine Biological Laboratory, Woods Hole, Massachusetts 02543

(Received 27 August 1996; accepted for publication 15 February 1997)

Many damselfishes perform a courtship behavior known as the signal jump, in which they rise in the water column and then rapidly swim downward while producing a pulsed sound. Pulsed sounds produced during the signal jump of the damselfish *Dascyllus albisella* were analyzed to determine whether they were correlated with the signal jump distance or speed, and how they changed with propagation. No statistical relationship was found between signal jump speed or distance with the number of pulses, pulse period, or change in the peak frequency of pulses in a call. If echoes were present in the sound, the change in echo delay would likely have been too small for damselfish to detect. Sounds attenuated with distance such that the signal-to-noise ratio decreased from 17–25 dB at 1 to 2 m to 5–10 dB at 11 to 12 m. It is unlikely that *D. albisella* can detect sounds at or beyond 11–12 m from the sound source, based on noise masking data from other fishes. Pulse period is least affected by propagation when compared to peak frequency, pulse duration, interpulse interval, and coefficient of variation of pulse amplitudes within a call. These results suggest that the sound produced during the signal jump functions over short distances and that the pulse period provides the most reliable basis for signal identification. © 1997 Acoustical Society of America. [S0001-4966(97)05706-8]

PACS numbers: 43.80.Ka, 43.80.Jz [FD]

INTRODUCTION

Damselfishes (*Pomacentridae*) are well-known for the male courtship display known as the signal jump, in which a male rises in the water column and then swims down rapidly while producing a pulsed sound (Fishelson, 1964; Myrberg, 1972; Spanier, 1979; Lobel and Mann, 1995). The signal jump involves tremendous amounts of rapid swimming (up to 3000 signal jumps per day for *Dascyllus albisella*), and it has been proposed to act as an indicator of male vigor for females choosing mates (Gronell, 1989; Knapp and Kovach, 1991; Mann and Lobel, 1995). Because the signal jump is associated with sound production, females may be able to determine male courtship rates by listening, without watching, leaving time for other activities like feeding. The courtship sounds vary in features such as pulse number and pulse rate (Lobel and Mann, 1995), which led us to ask whether these characteristics were related to swimming speed or distance during the signal jump.

A relationship between sound characteristics and swimming behavior during the signal jump could be produced by four mechanisms: (1) The number of pulses that could be produced in a signal jump of a given length could be limited so that longer signal jumps had more pulses. (2) There could be a correlation between the energy available for producing rapid pulsation and swimming speed. (3) The air–water interface is an almost perfect reflective surface that produces echoes detectable by fishes (Fay *et al.*, 1983). Information on the depth of the sound-producing fish (and thus swimming

distance and rate between pulses in sound) could theoretically be contained in changes in the delay of echoes from the surface as the fish swam down. (4) The volume of the swim-bladder could change with increasing depth, affecting the dominant frequency of the sound.

While the acoustic environment could provide information about the sound-producing fish, it could also affect the quality of the signal. Wind, temperature gradients, the ground, and foliage can restrict or enhance the distance over which signals can be used for communication in terrestrial environments (Marten and Marler, 1977; Wiley and Richards, 1978; Richards and Wiley, 1980; Brenowitz, 1982; Wells and Schwartz, 1982). Physical studies on sound propagation in shallow water (relative to wavelength) suggests that sound propagation is limited by sound frequency and water depth, and that the high reflectivity of the air–water interface could degrade signal qualities over short distances (Rogers and Cox, 1988; Forrest *et al.*, 1993). The little work that has been done on the propagation of fish sounds in shallow water suggests that they attenuate greatly over short distances. Playbacks of the tonal boatwhistle of the toadfish in 1-m deep water (*Opsanus tau*) decreased 18 dB over 5-m distance in 1-m water depth, theoretically restricting communication to several meters (Fine and Lenhardt, 1983). Natural (i.e., fish-produced) grunt sounds produced by the squirrelfishes *Myripristis violaceus* and *Myripristis pralinus* in 5-m depth water attenuated 10 dB over 40 cm to about 25-dB signal-to-noise ratio (SNR) (225-Hz octave band) (Horch and Salmon, 1973).

There have been no studies on the propagation of the sounds produced by damselfishes, and only one on the propagation of natural fish sounds in their natural environ-

^{a)}Current address: Department of Zoology, University of Maryland, College Park, MD 20742-4415. Electronic mail: mann@zool.umd.edu

ment (Horch and Salmon, 1973). The temporal patterning of fish sounds has been shown to be important in species discrimination of pulsed sounds, and has been hypothesized to be the most important factor in fish communication (Myrberg and Spies, 1980; Winn, 1964). Despite the potential importance of temporal patterning, there have been no measurements of the effect of propagation on call characteristics of pulsed fish sounds.

This paper examines the sound produced during the signal jump of the damselfish *Dascyllus albisella* with regard to: (1) how it is related to the swimming behavior during the signal jump; (2) how the characteristics of the sound vary with distance in their natural environment; and (3) how far the signal can likely be detected.

I. METHODS

A. Recordings

Recordings of sound production by *Dascyllus albisella* were made in April 1994 at Johnston Atoll (16° 44' N, 169° 31' W) while SCUBA diving using a SONY V-9 video camera coupled to a hydrophone (nominal flat response 10–3000 Hz; nominal calibration of -162 dB *re*: 1 V/ μ Pa) (BioAcoustics Inc., Woods Hole, MA), and with a SONY Professional Walkman (WMDC6) tape recorder with two hydrophones. The hydrophones were buoyed 0.5 m above the bottom by attaching a float to the end of the hydrophones and taping the cable to booms that were placed on the bottom. The water depth was 7 m. The hydrophones were placed so that the sound-producing fish were 1 to 2 m away.

B. Video analysis of signal jump

Video recordings of 15 calls made by four males were made in approximately 4-m water depth by a SCUBA diver holding the video camera or with the camera resting on the bottom. The fish were collected after recording and their standard length and weight were measured. The path of the signal jump was analyzed frame-by-frame (30 frames/s) by marking the position of the fish eye on an acetate sheet. The position of a stationary object in the same plane as the fish was also marked on the acetate to adjust for camera motion during the signal jump. The positions of the marks on the acetate were digitized, and the distance between marks was calculated based on the longest measurement of fish standard length from three different frames. From these data the average signal jump swimming speed and distance were calculated.

C. Signal analysis

The sounds produced during the signal jump were captured at 10 kHz using the computer program SIGNAL (Engineering Design, Belmont, MA). The number of pulses and pulse period were calculated using the detection algorithm described below, and regressed on the average signal jump swimming speed and signal jump distance.

Acoustic measurements were made using an automatic detection algorithm, as opposed to manual, to determine on- and off times of pulses to avoid user bias. The following

steps in signal analysis are illustrated in Fig. 1. Signals were low-pass filtered at 1000 Hz. The resulting signal was divided by its rms amplitude, full-wave rectified; an envelope function was calculated that followed the signal as it increased and calculated an exponential decay (with a 3-ms time constant) as it decreased. Then the signal was smoothed with a 3-ms rectangular window. Individual pulses were detected by gating the signal, which determines the on and off times of pulses by comparing the pulse amplitude to a threshold value. The threshold was dynamically defined for each signal as the maximum noise level in the first 20 ms plus 20% of the maximum signal amplitude (after rectification, envelope, and smoothing). From these detected signals, pulse number, pulse duration, interpulse interval (IPI), pulse period (number of ms per pulse), and coefficient of variation (CV) of pulse amplitude within a call, were calculated (see Fig. 1 for definitions). Peak frequency for each pulse of a call was calculated with an n -point FFT and a Hanning window (where n was the next highest power of 2 to the number of points in the digitized signal).

D. Analysis of sound propagation

Stereo recordings for propagation analyses were made in 7-m water depth. The position of one hydrophone was fixed 1 to 2 m from the sound-producing fish, while the other hydrophone was moved to the following distance relative to the fixed hydrophone: 0.02, 1, 2, 4, 8, and 10 m. After recording, the signals from both channels were captured simultaneously at 25 kHz using SIGNAL, and low-pass filtered at 1000 Hz. The signals from the fixed and roving hydrophones were analyzed as above to detect the pulses and measure pulse duration, IPI, pulse period, and CV of pulse amplitudes within calls. Cross-correlation analyses were performed on entire calls and pulses isolated from calls. The signal at the fixed (closer) hydrophone was used to determine the timing of the pulses, which were then isolated, including 4 ms before their on times and 14 ms after their off times.

Power spectra were calculated for each pulse in a call using an n -point FFT as above, but with a rectangular window. One-third octave band spectra were calculated from these power spectra. Attenuation was calculated at each of the distances between the fixed and roving hydrophones by averaging the one-third octave band spectra of all the pulses in a call and then averaging the call spectra. Attenuation was calculated as $20 \log(\text{roving hydrophone voltage}/\text{fixed hydrophone voltage})$. Noise level was measured using the captured signal up to 20 ms before the first pulse of the call. The signal-to-noise ratio (SNR) of the roving hydrophone was calculated for each distance by averaging the SNRs of each pulse in a call and then averaging the SNRs of the calls. SNR was calculated as $20 \log(\text{signal voltage}/\text{noise voltage})$. The signal voltage also includes noise that co-occurred with the signal.

E. Effect of signal-to-noise ratio (SNR) on analysis

The SNR of sounds could affect measurements of the on- and off times of pulses. If the error were constant, then it might be possible to correct the measurements depending on

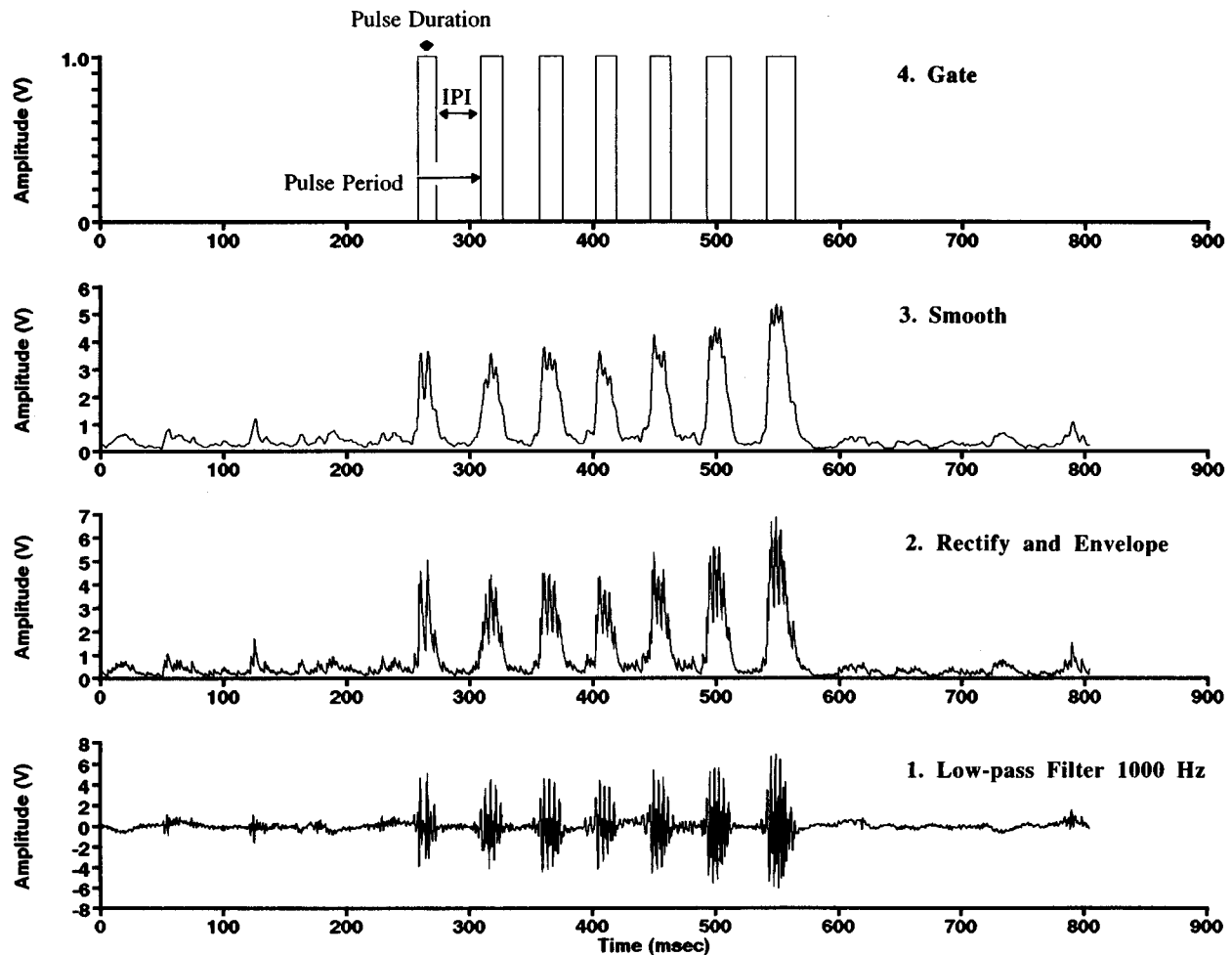


FIG. 1. Signal analysis of the sound produced during the signal jump. The original signal is shown in the bottom panel (panel 1) after low-pass filtering. The signal is then rectified and its envelope calculated with a 3-ms decay time (panel 2). After smoothing with a 3-ms window (panel 3), the signal is gated to determine on times (amplitude of 1) and off times (amplitude of 0) of the pulses. The characteristics measured from the gated signal are shown in panel 4.

the SNR. This was empirically tested by adding random noise to eight courtship signals with a high SNR, and measuring the change in pulse duration at different SNRs (Fig. 2). The coefficient of determination of the regression of SNR on the change in duration was 0.275, which indicated that it would not be possible to accurately determine pulse on- and off times using the SNR. However, there was less than 1-ms variation in pulse duration at SNRs greater than 19 dB. Therefore, only signals with SNR > 19 dB were used in the sound analysis of pulse duration and IPI.

II. RESULTS

A. Distribution of *Dascyllus albisella* at Johnston Atoll

Most *D. albisella* were found in the lagoon of Johnston Atoll, where depths are typically 3–5 m, except at locations that were dredged to depths of 7–15 m. The fish were less abundant on the outer reef slope, where they are found to depths of 30 m. They were rarely found near the reef edge, where the water is less than 1 m deep. The shallowest *D. albisella* were in 1.2-m water depth. While all recordings were made in situations where there were no coral heads

between the sound-producing fish and the hydrophone, there are many habitats where corals provide a natural obstruction to sound. In one area where 2 to 4-m-wide coral pinnacles rise from 6-m depth, *D. albisella* are located all around the coral heads and it would be impossible to make recordings 10 m away without an intervening coral head.

B. Signal jump analysis

The signal jump was characterized by initial, large, rapid beatings of the caudal fin and subsequent downward movement of the fish. Figure 3 shows a sample tracing of a signal jump and the frame-by-frame swimming speed. The signal jump swimming distance and speed were measured from the first large displacement to when the fish turned or the large displacements ended. Signal jumps averaged 80 cm/s (2.88 km/h), with sound produced over 57.7% of each jump on average (Table I).

Theoretically, if there was an echo from the surface during a signal jump starting from 1 m above the bottom, the average increase in the echo delay from the first to the last pulse (assuming the fish was directly above the hydrophone) would be 0.25 ms (Table I). The change in echo delay would be less if the fish was not directly above the hydrophone,

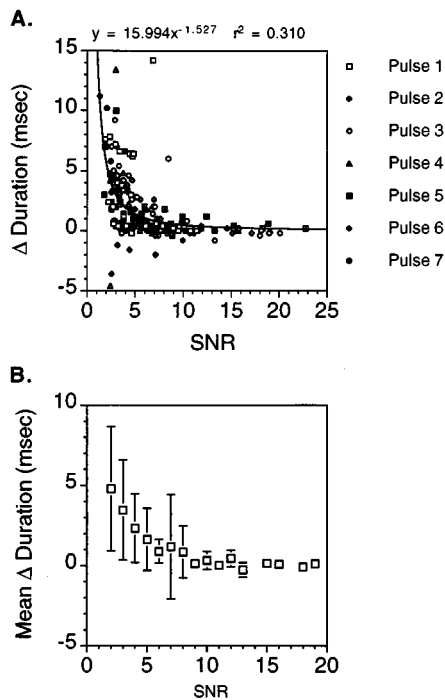


FIG. 2. Effect of SNR on the measured pulse duration (duration with no added noise duration with added noise). (A) Change in duration for each pulse in a call. (B) Mean (square) and 1 s.d. (error bars) of changes in duration for all of the pulses at each SNR.

which can be thought of as a model for a receiving fish. In deeper water the percent change in the echo delay would be less (Table I). If there was a surface echo, the duration of the pulses in a call should increase with each pulse, as the fish moves farther from the surface. Although 14 or 15 regressions of pulse duration on pulse number had positive slopes, none was statistically significant ($p > 0.05$) and the coefficients of determination were generally low (r^2 mean \pm s.d. = 0.24 ± 0.25 , range = 0.001–0.704). Furthermore, the variation in pulse durations within a call was much greater than that expected from an echo (mean \pm s.d. = 16.4 ± 3.0 ms, range = 12.0–24.7 ms).

The signal jump speed and signal jump length were positively correlated, meaning that fish swam faster during faster signal jumps ($r = 0.760$, $p = 0.0006$). However, the number of pulses and the pulse period of the sound produced during the signal jump were not related to either the signal jump speed or signal jump length (Fig. 4).

The peak frequencies of consecutive pulses in a call were analyzed to determine if the change in depth between pulses in a call produced a change in their peak frequency. The peak frequency of consecutive pulses did not vary regularly, and there was no statistically significant correlation between pulse number and peak frequency (r ranged from -0.773 to 0.884 ; $n = 15$ from four fish; only the two extremes of this range had $p < 0.05$). Similarly, the peak frequency of the sounds made during the signal jump were compared to those made during visiting and mating for five males in 4-m water depth; signal jumps are made higher in the water column than visiting and mating sounds which occur on the bottom (Mann, 1995). There was no significant

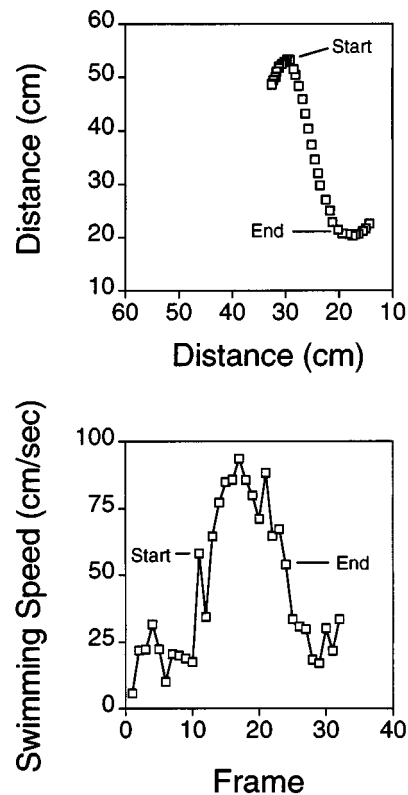


FIG. 3. Frame-by-frame position of a male *D. albisella* during a signal jump. Each point represents the position as traced onto acetate for a single frame. Each frame was separated by 33 ms. Start and End indicate the starting and ending points of the signal jump. The lower graph shows the swimming speed for each frame of the signal jump.

difference between these sound types and peak frequency ($p = 0.104$; two-way ANOVA with five males; $p = 0.615$ for difference in frequency between males).

C. Sound propagation

The effect of distance on sounds (including both courtship and aggressive sounds) was determined by comparing sounds received at two hydrophones separated by 0.02, 1, 2, 4, 8, and 10 m. Signals were evident up to hydrophone separations of 10 m (Fig. 5). Travel time between hydrophones was estimated by cross correlating the signals from the fixed and roving hydrophones using both whole calls [Fig. 6(A)] and pulses isolated from calls [Fig. 6(B)]. The sound speed was close to the predicted sound speed up to 4 m, while it varied more at 8 m and 10 m. The correlations between the signals at the two hydrophones decreased with distance for whole calls [Fig. 6(C)] and isolated pulses [Fig. 6(D)].

To determine the effect of distance on call characteristics, the differences between each signal at the fixed and roving hydrophone were calculated. Two results are important. First, whether there is systematic change in a call characteristic at different distances, and second, whether there is an increase in variance of a call characteristic with distance. The average peak frequency changed little over 8 m, but the variance increased; even at 2 m separation (Fig. 7). Pulse duration increased, and IPI and CV of pulse amplitudes decreased with distance, with most of the change in the first

TABLE I. Signal jump statistics and the potential effect of movement during sound production on echoes. Pulse duration mean and s.d. are calculated from means and s.d.'s of pulse durations within individual calls (excluding the first pulse).

| | Mean | s.d. | Minimum | Maximum | <i>n</i> |
|---|------|------|---------|---------|----------|
| Signal jump signal (cm/s) | 80 | 20 | 50 | 120 | 15 |
| Signal jump distance (cm) | 34 | 10 | 14 | 56 | 15 |
| Distance swam while making sound (cm) | 19 | 8 | 11 | 35 | 15 |
| Percent of jump making sound | 57.5 | 16.1 | 41.1 | 88.0 | 15 |
| Pulse duration (ms) | 16.4 | 3.0 | 12.0 | 24.7 | 15 |
| Maximum change in distance echo travel (cm) | 38 | | 22 | 72 | |
| Maximum change in time of echo delay (ms) | 0.25 | | 0.15 | 0.48 | |
| Echo delay in 4-m depth (ms) | 3.9 | | 3.9 | 3.9 | |
| Echo delay in 7-m depth (ms) | 7.8 | | 7.8 | 7.8 | |
| Percent change in 4-m depth | 6.5 | | 3.7 | 11.6 | |
| Percent change in 7-m depth | 3.2 | | 1.9 | 5.8 | |

2 m (Fig. 7). Pulse period was least affected by distance, with no significant difference between the signals received at the fixed and roving hydrophones. Furthermore, pulse period varied by only 2% compared to its mean value, whereas the IPI varied by about 10% of its mean value and the other parameters by 40% or more of their mean values (Fig. 7). There are no data at hydrophone separations of 10 m for pulse duration and IPI, because they did not have a SNR > 19 dB. There are also few data for pulse period and CV amplitude at 10 m, because many pulses that were detected at the fixed hydrophone were not strong enough at the roving hydrophone to be detected by the signal processing algorithm.

To understand how the variance in signal characteristics with propagation compares to the natural level of variation in the population, the variances in the call characteristics at 2-m

hydrophone separation were compared with population level variation measured from different individuals (data from Mann, 1995; data are from recordings made with the hydrophone 1 m above the bottom, 1 to 2 m from the sound-producing fish, data are from more than 100 calls) (Table II). This measure of population level variation will include some variation due to propagation effects, but it provides a useful measure for comparison. The variation in IPI and CV of pulse amplitudes at hydrophone separations of 2 m was greater than 60% of the natural population variation. There was more variation in pulse durations over 2 m for the same signals than there was natural variation in the population (Table II). The variation in pulse period due to propagation effects was only 10.3% of the population-level variation, thus the natural variation in the population would not be masked by the effects of propagation over short distances. Frequency varies for different sized males, the difference between the smallest and largest being about 160 Hz (Lobel and Mann, 1995). The standard deviation in the frequency of the same signals received 2-m apart was 153 Hz (Fig. 7).

Signals were most energetic between the 251- and 501-Hz one-third-octave bands (Fig. 8). The SNR in these bands decreased from 17–25 dB at 0-m hydrophone separation to 5–10 dB at 10-m hydrophone separation (Fig. 9). Attenuation did not vary regularly with frequency band, but generally decreased with distance (Figs. 8 and 9). Of the most energetic frequency bands, the lower frequency bands (251 and 316 Hz) attenuated up to 10 m, while the higher frequency bands (398 and 501 Hz) attenuated most in the first 4 m, with little attenuation afterward.

III. DISCUSSION

Since male damselfishes guard the eggs laid by females, there may be selective pressures for females to choose mates that will best defend their clutch. The repetition rate of signal jumps may act as an indicator of male vigor (Knapp and Kovach, 1991; Karino, 1995; Mann and Lobel, 1995). The speed and distance of signal jumps could also reliably indicate male quality because of the amount of energy invested in swimming. The first part of this study investigated when

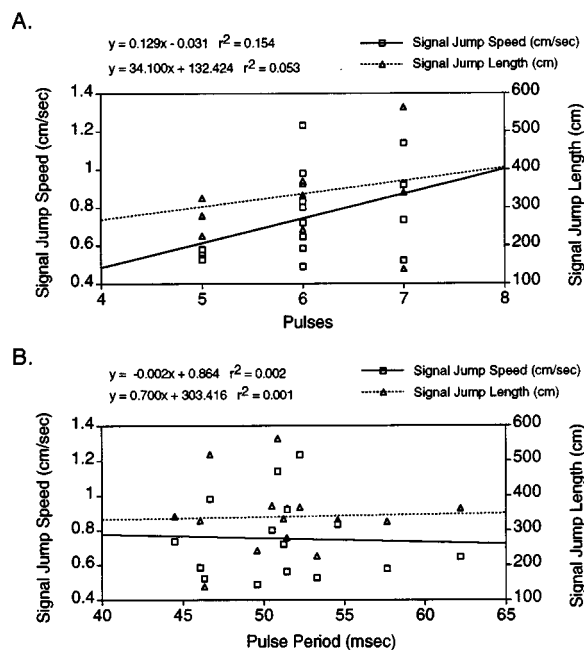


FIG. 4. (A) Regression of the number of pulses in a sound on signal jump speed (squares) and signal jump length (triangles; second y axis). (B) Regression of pulse period on signal jump speed and signal jump length.

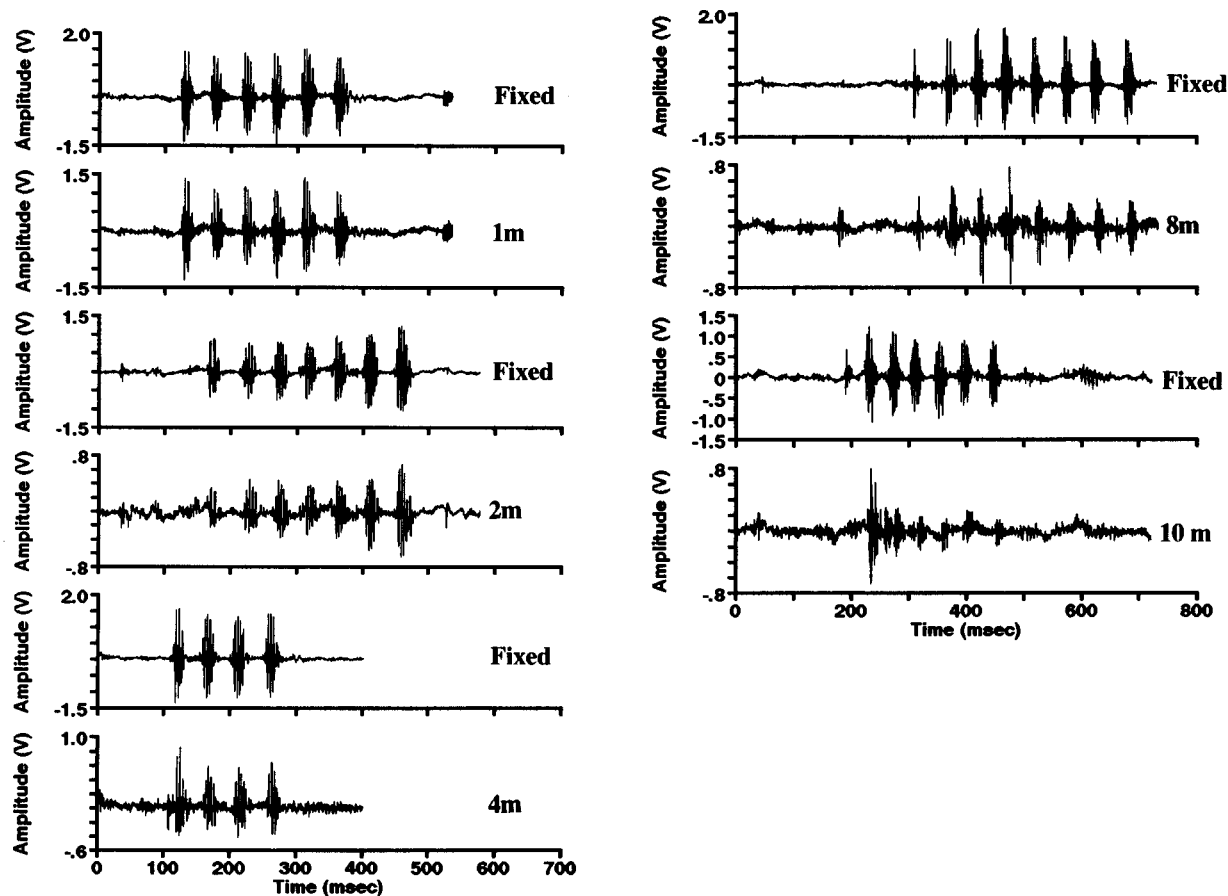


FIG. 5. Oscillograms of the calls from the fixed and roving hydrophones. The distances indicate the separation between the fixed and roving hydrophones. Each pair of signals is a separate call.

sound was produced during the signal jump and whether sound characteristics, including pulse number, pulse rate, and change in pulse frequency, were correlated to the signal jump swimming speed or distance. The second part of this study examined how sound characteristics were affected by sound propagation.

A. Pulse number, pulse rate, and frequency

Pulse number and pulse rate could not be used to accurately predict signal jump speeds or distances. Pulse number and pulse rate, as we discuss below, are least affected by signal propagation and thus would be reliable carriers of information about the signal jump. However, the variation in pulse number and pulse rate carries no information about variation in the signal jump.

The peak frequency of the sound could be affected both by the behavior of the fish and by the environment. The resonance frequency of the swimbladder might be expected to increase as it decreased in volume during the dip. Yet the distance traveled by a fish while producing the sound (19 cm on average) would yield only a very small change in swimbladder volume (1.5% for a 19-cm dip starting at 3-m depth). The variation in frequency between pulses might be the result of propagation effects, or due to changes in the tension on the swimbladder produced by the swimming fish.

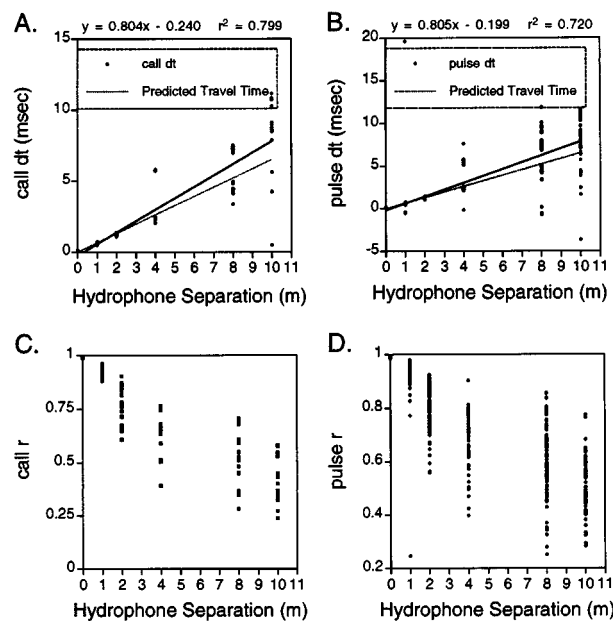


FIG. 6. (A) and (B) Travel time between the fixed and roving hydrophones by the amount of hydrophone separation. Travel times are the lags from a cross correlation of the signals [either calls (panel A) or individual pulses (panel B)] received at the fixed and roving hydrophones that produced the maximum positive correlation. (C) and (D) The maximum positive correlations from the autocorrelations of calls [panel (C)] and pulses [panel (D)].

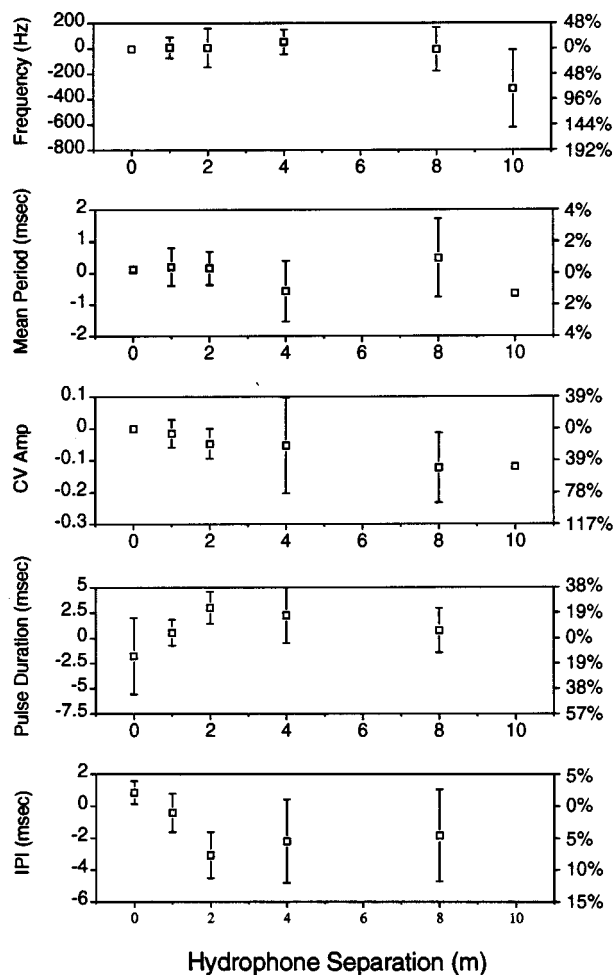


FIG. 7. Mean and 1 s.d. (first y axis) of the difference in call characteristics at the roving hydrophone minus the fixed hydrophone by the distance of separation. Call characteristics are peak frequency, pulse period, CV of pulse amplitudes within a call, pulse duration, and interpulse interval (IPI). The second y axis is the percent deviation from the average value of the parameter.

B. Repetition noise

Changes in the time of arrival of echoes produce a frequency sensation in humans known as repetition noise. Goldfish are able to detect the presence of repetition noise with delays ranging from 1.5 to 10 ms, corresponding to a reflecting surface 1–7.5 m away (Fay *et al.*, 1983). Because the damselfish swim while producing sound, we hypothesized that changes in the delay of echoes may provide information on swimming distances and speeds.

TABLE II. Comparison of signal variation due to propagation effects (measured by differences in signals received at hydrophones separated by 2 m) versus natural variation in signals (population standard deviation). Data for population standard deviation from Mann (1995).

| Signal characteristic | Propagation effect (mean difference + s.d.) | Population standard deviation | Relative propagation effect to population s.d. |
|-----------------------|---|-------------------------------|--|
| Period (ms) | 0.7 | 6.6 | 10.6% |
| CV amplitude | 0.09 | 0.15 | 60.0% |
| Pulse duration (ms) | 4.6 | 3.2 | 143.8% |
| IPI (ms) | 4.5 | 6.6 | 68.2% |

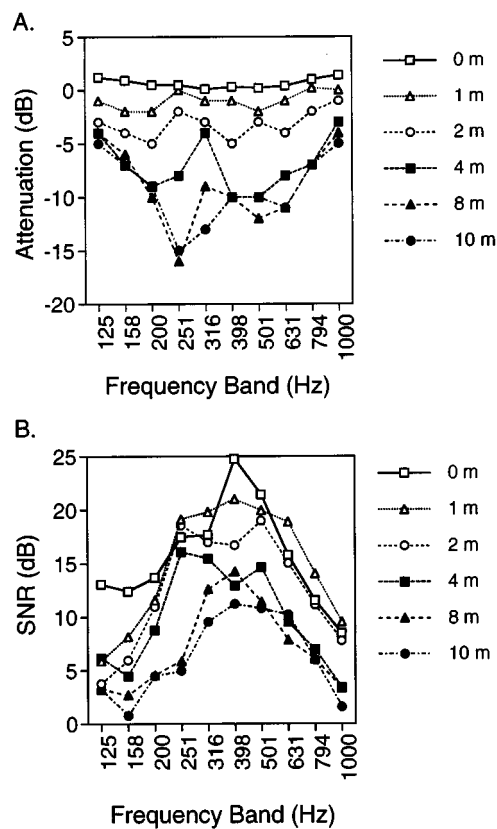


FIG. 8. One-third octave band spectra of average (A) Attenuation; and (B) SNR at the roving hydrophone.

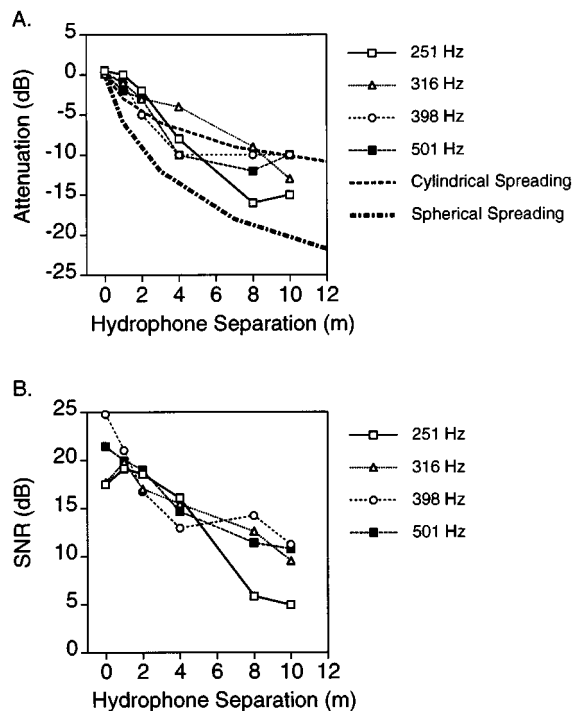


FIG. 9. (A) Attenuation with distance for the four most energetic one-third octave bands. Theoretical attenuation based on cylindrical and spherical spreading are shown by heavy dotted lines. (B) SNR of the signal at the roving hydrophone for the four most energetic one-third octave bands.

However, we could find little evidence for surface echoes in the damselfish calls. As the distance from the source to the receiver increases, the echo delay should decrease, yet the duration of pulses increased from 1 to 2 m away from the fixed hydrophone. Furthermore, surface echoes could not be consistently confirmed in analyses of individual pulses through autocorrelations and spectral analyses. Autocorrelations should show a negative correlation at the echo delay, which should also appear as a null in the power spectra at 1/echo delay (due to a 180-deg phase shift in the echo).

Assuming that there was an echo, the change in echo delay would have been about 6.5% from the first to the last pulse in 4-m water depth, and 3.2% in 7-m water depth. Goldfish (*Carassius auratus*) can detect 6% changes in echo delay when the echo is the same intensity as the original signal, but only 20% changes when the echo is attenuated 15 dB relative to the original signal (Fay *et al.*, 1983). If the detection abilities of damselfishes are the same as goldfish, then the differences in echo delay from the first to the last pulse would be barely detectable or undetectable. It is important to note, however, that goldfish (*Otophysi: Cyprinidae*) have specializations for sound detection that damselfish lack (Popper, 1982).

If changes in delays are easier to detect at lower levels of echo attenuation, then these should occur far from the source, where the path of the echo is not much longer than the direct path. However, there will be less change in echo delay farther from the source. Thus there may be a narrow window over which changes in echoes are detectable by fish in these environments, which will be affected by the depth of the water, the depth of the source, and the depth of the receiver. More data are needed on the discrimination of changes in echo delays at different levels of attenuation, before this issue can be fully addressed.

C. Propagation

The second part of this study investigated the effects of propagation on signal quality. The propagation of sound in shallow water (where the wavelength of the sound is about the same or larger than the water depth) is complex. Propagation can drastically vary depending on the depth of the sound source and receiver (Forrest *et al.*, 1993). The variation in the signal characteristics due to propagation over these short distances was similar to the population variability for all characteristics except pulse period. The variation due to propagation effects would make it difficult to detect differences in signals varying in pulse duration, IPI, CV of amplitude, and dominant frequency. This supports Myrberg's (1980) prediction, and suggests that pulse period may encode most of the reliable information about a fish sound. Although the number of pulses detected by the signal processing methods decreased with distance, it is not known how well fish can detect the number of pulses.

Overall, the courtship sound attenuated between what would be expected for cylindrical and spherical spreading. The 398- and 501-Hz frequency bands approximated cylindrical spreading and showed little attenuation beyond 4 m. This could be due to an echo from the bottom. Since the hydrophones are 0.5 m off the bottom and fish usually call

1–3 m from the bottom, the angle of incidence is small at distances greater than 4 m. Depending on the sound speed of the bottom, below a critical angle much of the sound will be reflected (Rogers and Cox, 1988). The critical angle of a fine sand bottom (28 deg) would produce a bottom echo at distances greater than 4 m (for a fish calling 1 m above the bottom and a receiver 1 m above the bottom). Although the spreading data suggest the existence of a bottom echo, a known source would need to be used to verify its presence.

The two lower energetic frequency bands continued to attenuate beyond 4 m. This might be due to the cutoff frequency of the channel, below which frequencies do not propagate. In a 7-m channel the cutoff frequency ranges from 116 Hz (for a fine sand bottom) to 1070 Hz (for a sand–silt–clay bottom) (Rogers and Cox, 1988). The speed of sound in the bottom where our recordings were made is unknown; it was a complex mixture of carbonate sand, shells, live coral, and coral rubble.

Audiograms of fishes have shown that they can be greatly influenced by ambient noise, and that masking by background noise limits hearing sensitivity. In general, SNRs are 15–20 dB at thresholds of detection of pure tones (Buerkle, 1969; Chapman, 1973; Chapman and Sand, 1974; Fay, 1974; Hawkins and Sand, 1977; Fay, 1988). When there are directional differences between the noise sources and the sound source, the SNR at threshold decreases to 10–15 dB (Chapman, 1973). The only data on masking in damselfish are from *Stegastes partitus* in which the detection threshold is 14-dB SNR (Ha, 1973). At 11 to 12 m (hydrophone separations of 10 m) the SNR of *D. albisella* sounds was 11 dB. If their ability to detect sounds in noise is the same as other fishes, then it is unlikely that they could detect sounds at or beyond 11 to 12 m from the source.

D. albisella could possibly use the SNR and duration of pulses in a call to determine the range to the source. However, given the range over which these sounds are likely detectable (<11 to 12 m), a female could easily see the sound producer performing the signal jump.

IV. CONCLUSIONS

Sound is produced on average during 58% of the signal jump. We did not detect any feature of the sound that was a good estimator of the signal jump distance and speed. This suggests that the mechanism of sound generation is not linked to the motor patterns producing the signal jump.

The diversity of acoustic habitats in which *D. albisella* are found, from 30-m depth on a sloping bottom, to 1.2-m depth on a flat bottom, should select for sounds that provide a reliable signal in a wide variety of situations. It is unlikely that changes in echo delays play a major role in the courtship behavior of *D. albisella*, because they would be difficult to detect and would be virtually absent in some habitats (e.g., 30-m depth).

This study utilized two hydrophones at one depth to gain an understanding of how parameters of the sounds produced by *D. albisella* were affected by propagation, and the distance over which the signal is likely to be detectable. While more complex hydrophone arrays and known sound sources would be necessary to fully characterize the propagation of

these sounds, our results support predictions that damselfish sounds are used over distances less than 11 to 12 m, and that the most reliably propagating information is contained in the pulse period of the sound.

ACKNOWLEDGMENTS

We thank Diana Ma, Alistair Economakis, and Gorka Sancho for diving assistance in the field. Additional support at Johnson Atoll was provided by Gary McCloskey, G. W. Walton, Mary Boyce, David Shogren, and many others in the office of the Program Manager for Chemical Demilitarization and Roger DiRosa and others of the U.S. Fish and Wildlife Service. We also thank Steve Oliver for his help with the video analysis of damselfish courtship. Rob Fricke, John Buck, and Kim Beeman were extremely helpful in discussions of acoustic analyses. Peter Tyack, Rob Fricke, and Arthur Popper provided helpful comments on the manuscript. The research was supported by grants from the U.S. Army Chemical Material Destruction Agency (via NOAA Sea Grant No. NA90-AA-D-SG535 and the Office of Naval Research N00014-91-J1591 and N00014-92-J-1969), the U.S. Army Legacy Resource Management Program (DAMD 17-93-J-3052), the Island Foundation, and the Kelley Foundation.

- Brenowitz, E. A. (1982). "The active space of red-winged blackbird song," *J. Comp. Physiol. A* **147**, 511–522.
- Buerkle, U. (1969). "Auditory masking and the critical band in Atlantic Cod (*Gadus morhua*)," *J. Fish. Res. Bd. Can.* **26**, 1113–1119.
- Chapman, C. J. (1973). "Field studies of hearing in teleost fish," *Helv. Meer.* **24**, 371–390.
- Chapman, C. J., and Sand, O. (1974). "Field studies of hearing in two species of flatfish *Pleuronectes platessa* (L) and *Limanda limanda* (L) (Family Pleuronectidae)," *Comp. Biochem. Phys. A* **47**, 371–385.
- Fay, R. R. (1974). "Masking of tones by noise for the goldfish (*Carassius auratus*)," *J. Comp. Physiol. Psych.* **87**, 708–716.
- Fay, R. R. (1988). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL).
- Fay, R. R., Yost, W. A., and Coombs, S. (1983). "Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system," *Hearing Res.* **12**, 31–55.
- Fine, M. L., and Lenhardt, M. L. (1983). "Shallow-water propagation of the toadfish mating call," *Comp. Biochem. Physiol. A* **76**, 225–231.
- Fishelson, L. (1964). "Observation on the biology and behavior of Red Sea coral fishes. Contributions to the Knowledge of the Red Sea," **30**, 11–26.
- Forrest, T. G., Miller, G. L., and Zagar, J. R. (1993). "Sound propagation in shallow water: implications for acoustic communication by aquatic animals," *Bioacoustics* **4**, 259–270.
- Gronell, A. M. (1989). "Visiting behavior by females of the sexually

- dichromatic damselfish, *Chrysiptera cyanea* (Teleostei: Pomacentridae): a probable method of assessing male quality," *Ethology* **81**, 89–122.
- Ha, S. J. (1973). "Aspects of sound communication in the damselfish *Eupomacentrus partitus*," MS thesis, University of Miami.
- Hawkins, S. D., and Sand, O. (1977). "Directional hearing in the median vertical plane by the cod," *J. Comp. Physiol. A* **122**, 1–8.
- Horch, K., and Salmon, M. (1973). "Adaptations to the acoustic environment by the squirrelfishes *Myripristis violaceus* and *M. pralinius*," *Mar. Behav. Physiol.* **2**, 121–139.
- Karino, K. (1995). "Male-male competition and female mate choice through courtship display in the territorial damselfish *Stegastes nigricans*," *Ethology* **100**, 126–138.
- Knapp, R. A., and Kovach, J. T. (1991). "Courtship as an honest indicator of male parental quality in the bicolor damselfish, *Stegastes partitus*," *Behav. Ecol.* **2**, 295–300.
- Lobel, P. S., and Mann, D. A. (1995). "Spawning sounds of the damselfish *Dascyllus albisella* (Pomacentridae) and relationship to male size," *Bioacoustics* **6**, 187–198.
- Mann, D. A. (1995). "Bioacoustics and Reproductive Ecology of the Damselfish *Dascyllus albisella*," Ph.D. thesis, MIT-WHOI Joint Program in Biological Oceanography, Woods Hole, MA.
- Mann, D. A., and Lobel, P. S. (1995). "Passive acoustic detection of sounds produced by the damselfish *Dascyllus albisella* (Pomacentridae)," *Bioacoustics* **6**, 199–213.
- Marten, K., and Marler, P. (1977). "Sound transmission and its significance for animal vocalization," *Behav. Ecol. Sociobiol.* **2**, 271–290.
- Myrberg, Jr., A. A. (1972). "Ethology of the bicolor damselfish, *Eupomacentrus partitus* (Pisces: Pomacentridae): A comparative analysis of laboratory and field behavior," *Animal Beh. Mon.* **5**, 197–283.
- Myrberg, Jr., A. A. (1980). "Fish bio-acoustics: its relevance to the 'not so silent world,'" *Environ. Biol. Fish.* **5**, 297–304.
- Myrberg, Jr., A. A., and Spires, J. Y. (1980). "Hearing in damselfishes: An analysis of signal detection among closely related species," *J. Comp. Physiol.* **140**, 135–144.
- Popper, A. N. (1982). "The morphology and evolution of the ear in actinopterygian fishes," *Am. Zool.* **22**, 311–328.
- Richards, D. G., and Wiley, R. H. (1980). "Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication," *Am. Nat.* **115**, 381–399.
- Rogers, P. H., and Cox, M. (1988). "Underwater sound as a biological stimulus," in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga (Springer-Verlag, Berlin), pp. 131–149.
- Spanier, E. (1979). "Aspects of species recognition by sound in four species of damselfishes, genus *Eupomacentrus* (Pisces: Pomacentridae)," *Z. Tierpsychol.* **51**, 301–316.
- Wells, K. D., and Schwartz, J. J. (1982). "The effect of vegetation on the propagation of calls in the neotropical frog *Centrolenella fleischmanni*," *Herpetologica* **38**, 449–455.
- Wiley, R. H., and Richards, D. G. (1978). "Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations," *Behav. Ecol. Sociobiol.* **3**, 69–94.
- Winn, H. E. (1964). "The biological significance in fish sounds," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 213–231.

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.

Background contributions and coupling coefficients for backscattering by thick shells

Steven G. Kargl

Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Philip L. Marston

Department of Physics, Washington State University, Pullman, Washington 99164-2814

(Received 27 December 1995; revised 15 January 1997; accepted 29 January 1997)

Approximations for leaky wave coupling coefficients and background scattering contributions are usually derived by assuming the shell response is governed by thin shell theory. Comparisons are described with approximations derived by other methods for a thick spherical shell at frequencies above coincidence. The variation of the phase of the coupling coefficient from numerical evaluation of the Watson transformation residue by Kargl [Ph.D. thesis, Washington State University (1990)] is evident in an analytical approximation. The phase of the background includes inertial and wave-kinematic contributions and is related to research by Gaunaurd and Werby [J. Acoust. Soc. Am. **90**, 2536–2550 (1991)]. The inertial contribution to the background phase is derived directly from the reflectivity of a plate in the limit of small phase shift for propagation across the thickness of the shell. The frequency and angle dependence of the phase of the exact reflection coefficient of a plate (having the same thickness as the shell) is also examined. © 1997 Acoustical Society of America. [S0001-4966(97)02206-6]

PACS numbers: 43.20.Dk, 43.20.Fn, 43.40.Ey [JEG]

INTRODUCTION

Quantitative ray methods for backscattering by elastic shells in water require an understanding of both the coupling of the acoustic field with waves guided by the shell and the background contributions to the scattering associated with reflection and curvature. A common approach to the approximation of such contributions is to assume that for the frequency range of interest the dynamical response of the shell is governed by thin shell theory.^{1–3} Unfortunately, for many practical applications such as high-frequency (high-resolution) sonar systems, thin shell equations may not properly account for the elastic response of the scatterer. An alternate approach to approximating the required coupling coefficient does not directly use thin shell assumptions but compares the form of ray approximations with the form of resonance scattering theory (RST).^{4,5} The simplest version of that approximation⁴ applied to spherical shells gives a vanishing high-frequency limiting phase for the coupling coefficient G_l which approximates the launching and detachment of leaky waves. That approximation was found to give a useful ray synthesis of the backward⁶ and forward⁷ scattering by a thick shell over a range of frequencies.

Another approach to the evaluation of G_l for a thick spherical scatterer or a solid shell is the numerical evaluation of the residue associated with Watson transformation leaky wave poles. While for solid elastic spheres^{4,8} $|\arg(G_l)|$ was typically ≤ 0.15 rad, Kargl⁹ found that for a thick shell $|\arg(G_l)|$ tends to increase with decreasing frequency. An increase in $|\arg(G_l)|$ with decreasing frequency was found from thin shell theory³ as well as from the method of comparison with RST.⁵ A related variation in the phase of the leaky wave coupling coefficient for a circular cylinder was shown using a high-frequency convolution formulation.¹⁰ One of the purposes of this publication is to compare Kargl's numerical thick shell results⁹ for $\arg(G_l)$ with Marston's approximation.⁵ The numerical evaluation of the residue also shows that for the thick shell considered $|G_l|$ is well approximated by the standard result^{1,3,4} reviewed below in Eq. (9).

One approach to approximating the background contribution discussed by Norris and Rebinsky, Eqs. (41), (15), and (7) of Ref. 2, accounts only for the inertia of the shell and the spread of the geometrically reflected wavefront resulting from the curvature of the outer surface of the shell. Another approach¹¹ sums the ray contributions due to the

reverberation of longitudinal waves traversing repeatedly across the thickness of the shell. A third approach makes use of properties of a background^{12,13} that has been used for isolating shell resonances in the RST formulation.¹⁴ The relationship between these approaches and their physical significance is examined in the present paper.

In our discussion, the contributions to the complex form function f are related to contributions to the far-field scattered pressure p_s and to the incident pressure p_{inc} through the usual relationship $p_s = p_{inc}(fa/2r)\exp(ikr - i\omega t)$, where a is the radius of the sphere and r is the distance to the observer. Calculations are shown for a thick evacuated spherical shell of 440c stainless steel in water having a thickness $h = 0.162a$. The bulk material properties are longitudinal and shear velocities of $c_L = 5.854$ mm/ μ s and $c_s = 3.150$ mm/ μ s and density $\rho_E = 7.84$ g/cm³. The surrounding water has $c = 1.479$ mm/ μ s and $\rho = 1.00$ g/cm³. Leaky wave properties for this shell from the Watson methodology are discussed in Refs. 6 and 7 and are supported by scattering measurements.^{7,15}

I. PHYSICAL ASPECTS OF THE BACKGROUND CONTRIBUTION

The background contribution to f is conveniently expressed in terms of a complex effective reflection coefficient R as follows where $x = ka$,

$$f_b = R e^{-i2x}. \quad (1)$$

The form of (1) is motivated by the well-known result¹⁶ that for reflection by a fixed rigid sphere $R \rightarrow 1$ as $x \rightarrow \infty$. The approximation of Norris and Rebinsky³ may be put in the form of (1) by setting $R = R_N$ with

$$R_N = (x_N + ix)/(-x_N + ix) = \exp(i\phi_N), \quad (2a)$$

$$x_N = \rho a / \rho_E h, \quad (2b)$$

$$\phi_N = -2 \tan^{-1}(\rho / kh \rho_E), \quad (2c)$$

where x_N is the dimensionless null frequency. The simplest thick shell background contribution^{6,11} is given by taking R to be the complex unimodular reflection coefficient of a vacuum-backed plate at normal incidence

$$R_{KM} = \frac{(1 - \rho^2) \exp(2ik_L h)}{1 - \rho \exp(2ik_L h)}, \quad (3a)$$

$$\rho = (\rho_E c_L - \rho c) / (\rho_E c_L + \rho c), \quad (3b)$$

where $k_L = \omega / c_L = kc / c_L$ and ρ is the reflection coefficient of an elastic half-space having the same material as the shell. Kargl and Marston^{6,11} have also calculated geometrically a curvature correction to the specular contribution that is predicted to be appreciable when x is either small or lies close to the frequency of a longitudinal resonance. That resonance first occurs at $x_{LR} = \pi a c_L / ch \approx 76.8$. For the shell under consideration, the lower bound for neglecting the correction is $x \approx 6$. For simplicity we restrict our discussions to regions of x where the curvature correction (given by a ray series that is not trivially summed) may be neglected.

Comparison of Eqs. (2) and (3) shows that while R_N does not depend on c_L , R_{KM} does through both $k_L h$ and ρ .

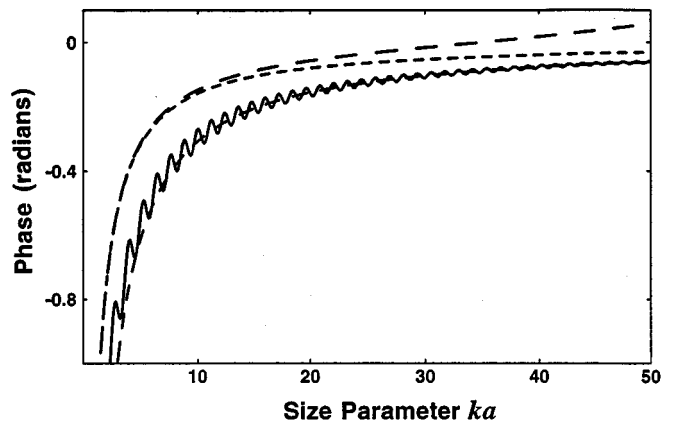


FIG. 1. Comparison of the phase of the effective reflectivity R in Eq. (1) for different background approximations for a spherical shell with a thickness to radius ratio $h/a = 0.162$. The upper curve with the longest dashes is the elastic plate approximation, Eq. (3), which merges with the adjacent curve ϕ_N from Eq. (2) at low frequencies. The solid curve is from the background given by Werby and Gaunard (Refs. 12 and 13). The adjacent dashed curve is $\phi_N + \phi_{LK}$ where ϕ_{LK} is the leading order wave-kinematic correction for a rigid sphere, Eq. (8). For ka of 10 the phase shift $k_L h$ for propagation once across the thickness of the shell is 0.41. Appendix A shows that ϕ_N is recovered from Eq. (3) for small $k_L h$.

When $2k_L h \ll 1$, the more general expression R_{KM} reduces to the R_N as shown in Appendix A: R_N includes only inertia but not any elastic effects on reflection off the shell. The equivalence of these approximations at sufficiently low frequencies is shown in Fig. 1 where the curve with long dashes gives $\phi_{KM} = \arg(R_{KM})$ and the adjacent curve with short dashes is ϕ_N . Since $x_N = 0.79$ in this example, the similarity between ϕ_N and ϕ_{KM} extends far above x_N . For certain applications, such as the interpretation of the bandlimited impulse response of thin shells, R_N has the computational advantage of having a simple Fourier transform (Appendix C of Ref. 17); R_N smoothes out unresolved features of the bandlimited response. For other applications, such as backscattering by thick shells it may be important to include the dependence on $k_L h$ retained in both R_{KM} and in the geometric curvature correction. Both ϕ_N and ϕ_{KM} display a transition from the phase of a soft reflector as $x/x_N \rightarrow 0$ to a hard reflector as x increases above x_N though ϕ_{KM} becomes positive at sufficiently large x .

The RST approach¹⁴ expresses the background as a series given by an infinite sum of partial waves of the form $f_n^{(b)} = -i(-1)^n(2n+1)[S_n^{(b)}(x) - 1]/x$, where $S_n^{(b)}(x)$ is known as the unimodular S function for the n th partial wave. Consider the following choice⁵

$$S_n^{(b)} = - \frac{[h_n^{(2)'}(x) - B(x)h_n^{(2)}(x)]}{[h_n^{(1)'}(x) - B(x)h_n^{(1)}(x)]}, \quad (4)$$

where $h_n^{(1,2)}$ are spherical Hankel functions and the prime denotes differentiation with respect to x . The real background transition function $B(x)$ is such that $B = 0$ and ∞ give $f_n^{(b)}(x)$ to be that of a rigid and soft scatterer, respectively. The choice

$$B(x) = 3\rho x / \{\rho_E(1 + x^2)[1 - (b/a)^3]\}, \quad (5)$$

recovers the Werby and Gaunard approximation^{12,13} of $f_n^{(b)}$ which when summed gives the background form function f_{WG} . It is convenient to introduce a complex effective reflectivity $R_{WG} = f_{WG} \exp(i2x)$.

Similarities and differences between R_{WG} and R_N will now be examined for the thick shell under consideration. Since $|R_{WG}|$ is found to differ from unity by less than 0.06 for $x \geq 10$, attention will be restricted to the phase. To simplify the comparison, (5) is replaced by

$$B(x) \approx (\rho a / \rho_E h) [x / (1 + x^2)] \quad (6)$$

since h/a is somewhat smaller than unity. The solid curve in Fig. 1 shows $\phi_{WG} = \arg(R_{WG})$ which is offset from ϕ_N , the curve with the shortest dashes. The offset is because f_N neglects a wave-kinematic background term (due to the curvature of the reflecting surface) that can be important when x is not very large. A Luneburg–Kline asymptotic expansion in inverse powers of x yields the following effective reflectivity for a fixed-rigid reflector¹⁶

$$R_{LK} \approx 1 - (3i/2x) - (5/2x^2). \quad (7)$$

The amplitude $(R_{LK} - 1)$ may be introduced as a correction term to the background f_N to leading order in $1/x$ by introducing the phase shift

$$\phi_{LK} = -\tan^{-1}(3/2x), \quad (8)$$

which is the phase of the first two terms in (7). The modified approximation is given by (1) with $R = \exp(i\phi_N + i\phi_{LK})$. The curve with the intermediate dashes in Fig. 1 shows $\phi_N + \phi_{LK}$. It closely follows the average trend in ϕ_{WG} when $x \geq 10$. While the $B(x)$ used in this comparison will gain additional significance from results discussed in Sec. II, the comparison suggests that ϕ_{LK} can give a significant contribution to $\arg(R)$ for thick shells.

Comparisons like those shown in Fig. 1 were also computed for thinner shells of the same material having $h/a = 0.05$ and 0.025 . In each case the difference between ϕ_N and ϕ_{KM} was much smaller than for Fig. 1. Furthermore $\phi_N + \phi_{LK}$ reproduced the trend in ϕ_{WG} for $x \geq 10$ and the fine oscillations in ϕ_{WG} were suppressed. A much larger fraction of the total phase was from ϕ_N suggesting that for thin shells the inertial correction dominates the phase of the background.

If the frequency is sufficiently large and/or the shell is sufficiently thick that ϕ_{KM} differs significantly from ϕ_N , wave-kinematic corrections may be introduced into the plate reverberation model by setting $R = \exp(i\phi_{KM} + i\phi_{LK})$. This is the approximation used in Sec. III.

II. COUPLING COEFFICIENT FOR THICK SHELLS

Let G_l denote the leaky wave complex coupling coefficient for spherical shells defined in the usual notation.^{3,4,8} Kargl⁹ numerically evaluated Watson transformation leaky wave residues to obtain G_l as summarized in Appendix B. For brevity restrict attention to the $l = a_0$ and s_0 leaky Lamb waves. (The threshold for propagation of the a_1 Lamb wave is^{6,7} $x = 41$.) From the analysis given in Refs. 4 and 5, the anticipated form of G_l is

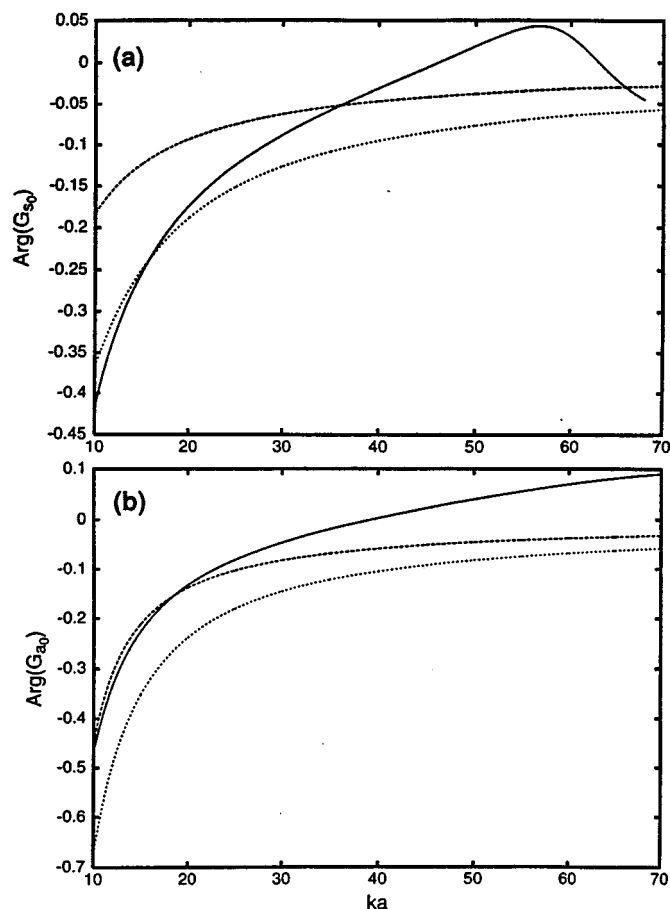


FIG. 2. Comparison of the phases of numerical coupling coefficients (solid curves) previously calculated by Kargl (Ref. 9) with Marston's approximations from Ref. 5 (dashed curves) evaluated for the (a) s_0 and (b) a_0 leaky Lamb waves on a shell with $h/a = 0.162$. The curve with the shortest dashes uses Eq. (6) to approximate the background transition function $B(ka)$ while the other dashed curve is for a rigid background which corresponds to $B = 0$ as discussed in Appendix C.

$$G_l = 8\pi\beta_l(c/c_l)e^{i\phi_l}, \quad (9)$$

where β_l and c_l denote the leaky wave radiation damping and phase velocity on the sphere which is assumed to be elastic. The leaky wave parameters are evaluated from the complex partial-wave index roots $\nu_l = \alpha_l + i\beta_l$ as described in Ref. 15. The analysis leading to the form of (9) assumes that^{4,5} $x \gg 1$ and $\alpha_l \gg \beta_l$. Numerical results (Appendix B) support the use of $|G_l|$ from (9). The error is greatest for the a_0 wave but is always less than 5% for $x > 10$ and it decreases rapidly with increasing x . For $l = s_0$ the value of β_l is small (typically less than 10^{-4}) for x near 21 causing $|G_l|$ to be small. (See Figs. 7 and 13 of Ref. 15. The extremely small value of β_l for this highly supersonic wave is attributed to a mode shape within the shell having almost no radial motion at the shell's outer surface.) The numerically determined ϕ_l are shown by the solid curves in Fig. 2(a) and (b). The trend of ϕ_l becoming more negative for decreasing x is consistent with the following thick shell approximation.

A relationship between ϕ_l and the $B(x)$ used in (4) was derived in Ref. 5 in which the residue of a complex frequency plane pole from RST is matched to the corresponding residue in the ray formulation. Phase information was re-

tained by using Debye approximations of Hankel functions while restricting attention to leaky waves (which have $c_l > c$). The result yields (9) with $\phi_l = \phi_1 + \phi_2 + \phi_3$ where⁵

$$\phi_1 = -2 \tan^{-1} \left\{ \frac{B \cos \theta_l}{\cos^2 \theta_l + (q/x)B + (q/x)^2} \right\}, \quad (10a)$$

$$\phi_2 = 2 \tan^{-1} \{ [3 + 5(\tan \theta_l)^2] / (24x \cos \theta_l) \}, \quad (10b)$$

$$\phi_3 = -2 \tan^{-1} \{ [1 + (\cos \theta_l)^{-2}] / (2x \cos \theta_l) \}, \quad (10c)$$

and $q = [1 + (\cos \theta_l)^{-2}] / 2$. The dashed curves in Fig. 2(a) and (b) show ϕ_l for two choices for B . The curves with the longer dashes are for $B=0$ which corresponds to a rigid background. In that case $\phi_l = \phi_2 + \phi_3 \approx -2 \tan^{-1} [(21 + 7 \tan^2 \theta_l) / (24x \cos \theta_l)] + O(1/x^2)$ which corresponds to a wave-kinematic contribution to ϕ_l given by Ho [Eqs. (9) and (12a) of Ref. 1]. The curves with the shorter dashes give ϕ_l where in ϕ_1 , B is given by (6). With that choice, through $O(1/x)$,

$$\begin{aligned} \phi_1 &\approx -2 \tan^{-1} (\rho a / \rho_E h x \cos \theta_l) \\ &= -2 \tan^{-1} (\rho / \rho_E k h \cos \theta_l), \end{aligned} \quad (11)$$

in agreement with a result derivable from Ref. 3 [Eqs. (4)–(6)]. Equation (11) has the form of a correction to $\arg(G_l)$ for cylindrical shells derived by a convolution method [Eq. (20) of Ref. 10] and indicates that ϕ_1 decreases as kh increases.

Inspection of Fig. 2 shows that the quality of the agreement between the approximations for ϕ_l and the numerical results (the solid curves) depends on the region of x and on the leaky wave type. For x sufficiently large that ϕ_{KM} differs significantly from ϕ_N in Fig. 1, it is to be expected that the function $B(x)$ in (6) does not capture all of the slowly varying features of the background. Consequently some differences between the solid curves and the sum of (10) are to be expected for $x \geq 30$. For x between 30 and 10, the difference between the solid and the short dashed curves is somewhat larger for $l=a_0$ than for $l=s_0$. Inspection of Fig. 2(b) suggests that for $l=s_0$ in (10a), B may be approximated by (6), but for $l=a_0$ it is more accurate to use $B=0$. This is confirmed in Appendix C by analyzing the angular evolution of the phase of the reflection coefficient for a plate having the same thickness as the shell. A relationship between the phase evolution and the ϕ_1 contribution to ϕ_l from Ref. 10 is used. Appendix C also examines how, for large x , the solid curves in Fig. 2 may be influenced by the onset of other Lamb waves.

The applicability of the Debye approximations used in the derivation of (7) breaks down if $(c_l/c - 1)$ and x are not sufficiently large.¹⁸ For $l=a_0$, c_l/c decreases to 1.25 as x decreases to 10 while for $l=s_0$, $c_l/c \geq 3.8$ in this region. According to Ref. 18 [Fig. 3 and Eq. (7)], the Debye approximation begins breaking down for the a_0 wave near the left side of Fig. 2(b). Physically this is associated with the overlap of the Fresnel width of the coupling region with the edge of the scatterer.¹⁸ It is noteworthy that the value of $|G_l|$ is approximated by $8\pi\beta_l c/c_l$ even if c_l/c is too small for the Debye approximations to be applicable. This follows from the factorization used in the derivation by comparison

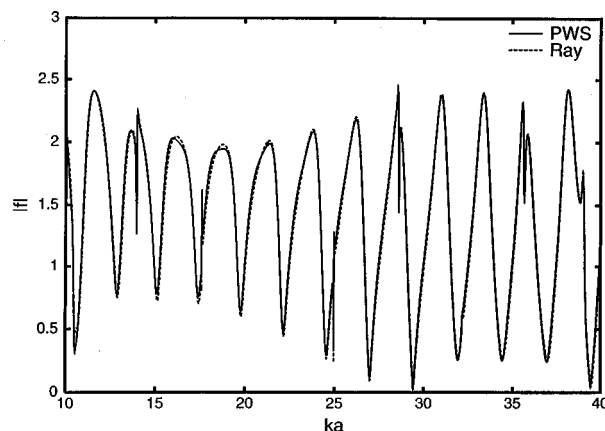


FIG. 3. Comparison of $|f|$ from the partial wave series from elasticity theory for a thick shell (solid curve) with the synthesis from Eq. (12) (dashed curve).

with $RST^{4,5}$ and because $S_n^{(b)}$ is always unimodular. Hence the result is also applicable to creeping and trapped waves (having¹ $c_l \approx c$ and $c_l < c$) provided $\alpha_l \gg \beta_l$.

III. SYNTHESIS FOR BACKSCATTERING BY A THICK SHELL AND DISCUSSION

Combining the approximations discussed in Sec. I and II with the previous result⁴⁻⁹ for the leaky wave contributions f_l gives the synthesis

$$f \approx R_{KM} e^{i(\phi_{LK} - 2x)} + f_{s_0} + f_{a_0}, \quad (12)$$

$$f_l = \frac{-G_l \exp[-2(\pi - \theta_l)\beta_l + i\eta_l]}{1 + \exp(-2\pi\beta_l + i2\pi xc/c_l)}, \quad (13)$$

and $\eta_l = 2x[(c/c_l)(\pi - \theta_l) - \cos \theta_l] - \pi/2$. The contributions from subsonic waves are neglected since x is assumed to be sufficiently greater than the coincidence frequency ($x \approx 6$ for the shell under consideration) that such waves should be significantly damped. In the approximation of $\arg(G_l)$, (10) is used with B given by (6) for $l=s_0$ and $B=0$ for $l=a_0$ as explained in Appendix C. The comparison of $|f|$ with the exact result is shown in Fig. 3. These results may be compared with our previous syntheses for the same shell shown in Ref. 6 (Figs. 4 and 9). The changes in the ray synthesis are small resulting from the improved approximations of the phases of the G_l and of the background contribution. Some improved agreement is evident near the left side of Fig. 3. For reasons noted in Appendix C, if accurate approximations to $\arg(G_l)$ for $l=a_0$ and s_0 are needed at larger x values than displayed in Fig. 3, the influence of other Lamb modes may need to be included.

Both ϕ_{LK} and $(\phi_2 + \phi_3)$ may be considered wave-kinematic corrections due to distinctions between high-frequency approximations of phase shifts and local solutions of the Helmholtz equation.¹⁶ Both ϕ_{LK} and $(\phi_2 + \phi_3)$ may be rewritten in a form that manifests a dependence on the distance of the virtual source of outgoing waves and the reflecting or radiating region of the shell's surface. The distance between the surface region radiating leaky waves and the relevant virtual caustic⁸ is $a \cos \theta_l$. If c_l approaches c ,

that distance vanishes and $\phi_2 + \phi_3$ diverges when Debye asymptotics are used as in the derivation of (10), the effect of the nonrigid background being approximated by (10a).

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research.

APPENDIX A: INERTIA DOMINATED LIMIT OF THE PLATE REFLECTION COEFFICIENT

Introduce the parameters $\eta = kh\rho_E/\rho = x/x_N$, $\zeta = \rho c/\rho_E c_L$, and $\epsilon = 2k_L h = 2\eta\zeta$. The inertia limited expression (2) becomes $R_N = -\exp(i2 \tan^{-1} \eta)$. Without approximation, the plate reflectivity from (3) may be written $R_{KM} = -\exp(i2\mu)$ where $\mu = (\epsilon/2) + \psi$ and

$$\tan \psi = (1 - \zeta)(\sin \epsilon) / [(1 + \zeta) - (1 - \zeta)\cos \epsilon] = \nu. \quad (\text{A1})$$

It follows that

$$\tan \mu = [\nu + \tan(\epsilon/2)] / [1 + \nu \tan(\epsilon/2)]. \quad (\text{A2})$$

We now show that (3) approaches (2) when ϵ is sufficiently small that only $O(\epsilon)$ terms are retained in expansions of (A1) and (A2). Expansions of (A1) and $\tan(\epsilon/2)$ give $\nu \approx (1 - \zeta)\epsilon/2\zeta + O(\epsilon^2)$, and $\tan(\epsilon/2) \approx \epsilon/2 + O(\epsilon^3)$. Insertion of these into (A2) gives $\mu \approx \tan^{-1}[\epsilon/2\zeta + O(\epsilon^2)]$ which is the desired result since $\eta = \epsilon/2\zeta$.

APPENDIX B: NUMERICAL WATSON TRANSFORM THICK SHELL RESIDUES

Each term of the partial wave series from elasticity theory contains a factor B_n/D_n in the notation of Refs. 4, 8, and 15 where B_n and D_n are 5×5 determinants since the shell is evacuated on the inside. Repeating the steps leading to (27)–(32) of Ref. 8 yields the relationship between G_l and the Watson transformation residue given by (14) of Ref. 4. Let $F_\nu(x) = -x_s^2 \rho D_\nu^{(1)}(x) / \rho_E D_\nu^{(2)}(x)$, where $x_s = xc/c_s$ and $D^{(1)}$ and $D^{(2)}$ are the cofactor expansions with respect to the elements d_{11} and d_{21} of D_n with n replaced by the complex partial wave index $\nu = \nu_l$. For evacuated shells $D^{(1)}$ and $D^{(2)}$ are 4×4 determinants. The functions $\mathcal{D}_\nu^+(x)$ and $\mathcal{D}_\nu^-(x)$ are defined from $F_\nu(x)$ and (10) and (11) of Ref. 4. The ν_l are given by roots of $\mathcal{D}_\nu^+(x) = 0$ and are also roots of $D_\nu(x) = 0$. The evaluation of the residue is facilitated by numerical evaluation of¹⁹

$$\left(\frac{d\mathcal{D}_\nu^+(x)}{d\nu} \right)_{\nu=\nu_l} = \frac{1}{2\pi i} \oint_C \frac{\mathcal{D}_\nu^+(x)}{(\nu - \nu_l)^2} d\nu. \quad (\text{B1})$$

The integral is over a counterclockwise closed contour C selected such that \mathcal{D}_ν^+ is analytic within and on C . The procedure was implemented and discussed in Appendix I of Ref. 9.

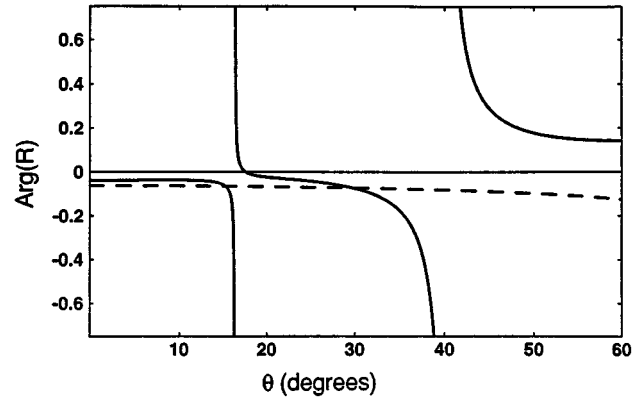


FIG. C1. The solid curve shows the phase (modulo 2π) of the unimodal reflection coefficient of a vacuum backed stainless-steel plate evaluated at $kh = xh/a$ with $x = 25$ and $h/a = 0.162$ and based on expressions given in Ref. 20. The full range of the vertical axis of $-\pi$ to π is not shown. The 2π phase jumps near 16° and 40° are associated with the s_0 and a_0 leaky Lamb waves. The background near each ramp gives a contribution to $\arg(G_l)$ (see text). The dashed curve is the approximation based on inertia, Eq. (C1).

APPENDIX C: PHASE OF THE THICK PLATE REFLECTION COEFFICIENT AS A FUNCTION OF ANGLE OF INCIDENCE AND RELEVANCE TO $\arg(G_l)$

Brekhovskikh²⁰ gives the exact expression for the phase ϕ_R of the unimodal reflection coefficient R of a vacuum-backed thick plate as a function of the angle of incidence θ . Figure C1 compares $\phi_R(\theta)$ for a plate having the same thickness as the shell under consideration with the following intertial approximation for thin plates or shells³

$$\phi_N(\theta) = -2 \tan^{-1}(\rho a / \rho_E h x \cos \theta), \quad (\text{C1})$$

(the dashed curve) which gives (2c) when $\theta = 0$. The approximation in (11) for ϕ_l , the background contribution to $\arg(G_l)$, is given by $\phi_N(\theta_l)$. The connection of $\arg(R)$ to ϕ_l follows from the product representation¹⁰ for R in the absence of dissipation:

$$R(k_p) = \frac{k_p - K_l^*}{k_p - K_l} e^{i\phi_b(k_p)}, \quad (\text{C2a})$$

$$= e^{i\phi_b(k_p)} + \frac{2iK_l''}{k_p - K_l} e^{i\phi_b(k_p)}, \quad (\text{C2b})$$

where $k_p = k \sin \theta$ is the projection of the incident wave number onto the plate, $K_l = K_l' + iK_l''$ is the location of the l th leaky wave pole in the complex k_p plane, and $\phi_b(k_p)$ varies slowly for k_p near K_l' . As a consequence of the symmetric pole-zero factor in (C2a), $\arg(R)$ undergoes an approximately 2π phase evolution as k_p increases from below K_l' to above K_l' . The corresponding transition in θ is from below $\theta_l = \sin^{-1}(K_l'/k)$ to above θ_l . The rate of evolution is faster when the leaky wave radiation damping rate K_l'' is small. The phase ramp near θ of $\theta_l = 16.36^\circ$ in Fig. C1 is from the s_0 leaky Lamb wave while the one near $\theta_l = 40.22^\circ$ is from the a_0 leaky Lamb wave. The phase ramp for the s_0 wave is narrow because of the corresponding pole and zero in (C2a) lie close to the real axis. The phase velocity ratios $c_l/c = k/K_l'$ of 3.55 and 1.55 for the s_0 and a_0

leaky plate Lamb waves are similar to the corresponding thick shell values of 3.84 and 1.68 tabulated⁹ from the Watson roots ν_l for $ka=25$. For $\theta \approx \theta_l$ for $l=s_0$ the phase variation is superposed on a slowly varying phase variation corresponding to $\phi_b(k_p)$ which may be approximated by the dashed curve from (C1). Similar comparisons show that (C1) tends to more accurately approximate the s_0 background phase for lower $x=ka$.

As demonstrated for the cylinders in Ref. 10, when $ka \gg 1$, $\phi_b(K'_l)$ gives a background contribution to $\arg(G_l)$ for the l th leaky wave. This is because each leaky wave scattering contribution is associated with the second term in (C2b). An analogous argument for the present sphere case indicates that at high frequencies, ϕ_1 , the background contribution to $\arg(G_l)$, may be approximated as $\phi_b(K'_l)$. Comparison of the exact and approximate results in Fig. C1 (and similar comparisons not shown here carried out for other values of ka) explains important differences in Fig. 2 between the s_0 and a_0 cases. Near θ_l for the s_0 pole, Eq. (C1) approximates the background phase offset except when ka is large (see below). It follows that (11) with B given by Eq. (6) should give a better approximation to ϕ_l than if a rigid background ($B=0$) is assumed. This is evident in Fig. 2(a) for $ka < 35$. Inspection of Fig. C1 (and similar plots for other ka values) shows that near θ_l for $l=a_0$, the background phase offset (modulo 2π) is smaller and is poorly approximated by (C1). This explains why the best approximation in Fig. 2(a) is with $B=0$ in (10). These assignments for B in (10) are confirmed by the synthesis shown in Fig. 3 and by a significant degradation of the synthesis (not shown) if, in Eq. (10), B from (6) is used for the a_0 wave. For ka above about 35, plots like Fig. C1 show that neither the s_0 or the a_0 wave has a background phase well approximated by (C1). This could have been anticipated since $\phi_{KM} = \arg(R_{KM})$ from (3) gives the exact intercept $\phi_R(0)$ and as shown by the uppermost curve in Fig. 1, ϕ_{KM} becomes positive above ka of 35. This may be attributed to the influence of pole-zero combination associated with the a_1 wave which has a propagation threshold at ka of 41. The form of (C2b) suggests that the propagation

threshold for each leaky wave affects $\arg(G_l)$ of the other waves. Equation (C2a) may be inverted to give $\phi_b(K'_l)$ for approximation of ϕ_1 for each G_l and ka of interest.

- ¹J. M. Ho, J. Acoust. Soc. Am. **95**, 1264–1273 (1994).
- ²A. N. Norris and D. A. Rebinsky, J. Acoust. Soc. Am. **95**, 1809–1829 (1994).
- ³D. A. Rebinsky and A. N. Norris, J. Acoust. Soc. Am. **98**, 2368–2371 (1995).
- ⁴P. L. Marston, J. Acoust. Soc. Am. **83**, 25–37 (1988). (See Secs. IV D and V D for application to shells.) See also P. L. Marston and K. L. Williams, J. Acoust. Soc. Am. Suppl. 1 **83**, S94 (1988).
- ⁵P. L. Marston, “Variable phase coupling coefficient for leaky waves on spheres and cylinders from resonance scattering theory,” Wave Motion **22**, 65–74 (1995).
- ⁶S. G. Kargl and P. L. Marston, J. Acoust. Soc. Am. **89**, 2545–2558 (1991).
- ⁷S. G. Kargl and P. L. Marston, J. Acoust. Soc. Am. **88**, 1103–1113 (1990).
- ⁸K. L. Williams and P. L. Marston, J. Acoust. Soc. Am. **78**, 1093–1102 (1985).
- ⁹S. G. Kargl, “Quantitative ray methods for scattering of sound by spherical shells,” Ph.D. thesis, Washington State University, 1990.
- ¹⁰P. L. Marston, J. Acoust. Soc. Am. **97**, 34–41 (1995). A minus sign was omitted from an expression like our (C1) given below (A2).
- ¹¹S. G. Kargl and P. L. Marston, J. Acoust. Soc. Am. **88**, 1114–1122 (1990).
- ¹²G. C. Gaunard and M. F. Werby, J. Acoust. Soc. Am. **90**, 2536–2550 (1991).
- ¹³M. F. Werby, J. Acoust. Soc. Am. **90**, 3279–3287 (1991).
- ¹⁴L. Flax, G. C. Gaunard, and H. Überall, in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1981), Vol. 15, pp. 191–194.
- ¹⁵S. G. Kargl and P. L. Marston, J. Acoust. Soc. Am. **85**, 1014–1028 (1989).
- ¹⁶J. B. Keller, R. M. Lewis, and B. D. Seckler, “Asymptotic solution of some diffraction problems,” Commun. Pure Appl. Math. **9**, 207–265 (1956).
- ¹⁷G. Kaduchak, C. S. Kwiatkowski, and P. L. Marston, J. Acoust. Soc. Am. **97**, 2699–2709 (1995).
- ¹⁸P. L. Marston, J. Acoust. Soc. Am. **96**, 1893–1898 (1994).
- ¹⁹P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953) p. 375.
- ²⁰L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980), 2nd ed., pp. 70–72; Eq. (10.9) where MZ_1 and NZ_1 are defined in Eq. (10.3) such that the dependence on Z_1 cancels.

Exact solution for one-dimensional acoustic fields in ducts with a quadratic mean temperature profile

B. Manoj Kumar and R. I. Sujith

Department of Aerospace Engineering, Indian Institute of Technology, Madras, India 600036

(Received 1 March 1996; revised 29 November 1996; accepted 19 December 1996)

The purpose of this Letter is to present an exact analytical solution for sound propagation in ducts with a quadratic mean temperature profile. Using appropriate transformations, the one-dimensional wave equation for ducts with an axial mean temperature gradient was reduced to the hypergeometric differential equation, whose solution can be expressed in terms of hypergeometric functions. The analysis neglects the effects of mean flow and is therefore valid only for small mean Mach numbers.

© 1997 Acoustical Society of America. [S0001-4966(97)04805-4]

PACS numbers: 43.20.Mv, 43.20.Bi [JEG]

LIST OF SYMBOLS

| | |
|------------|------------------------------------|
| a, b, c | constants in Eq. (4) |
| c_1, c_2 | constants in Eqs. (12) and (13) |
| P' | acoustic pressure |
| R | specific gas constant |
| s | transformation variable in Eq. (5) |
| t | time |
| T | temperature |
| x | distance |

| | |
|----------------------|--|
| α, β, ν | arguments of the hypergeometric function |
| ρ | density |
| ω | angular frequency |
| γ | $= C_P/C_V$, ratio of specific heats |
| ψ | digamma function |
| Γ | gamma function |
| superscripts | |
| ' | oscillating quantity |
| — | time-averaged quantity |

INTRODUCTION

The purpose of this Letter is to present an exact analytical solution for sound propagation in ducts with a quadratic mean temperature profile. The behavior of one-dimensional acoustic fields in ducts with an axial mean temperature gradient is a problem of considerable scientific and practical interest. There is a need to develop an understanding of the manner in which a mean axial temperature gradient (caused, for example, by heat transfer to or from the walls) affects the propagation of sound waves and the stability of small amplitude disturbances in a duct. Such understanding will improve existing capabilities for controlling combustion instabilities in propulsion and power generating systems, designing pulse combustors and automotive mufflers, analyzing the behavior of resonating thermal systems, analyzing thermo-acoustic engines, and measuring impedances of high-temperature systems (e.g., flames). Consequently, there exists a need for obtaining exact analytical solutions that describe one-dimensional wave systems in ducts with axial temperature gradients.

The behavior of one-dimensional waves in ducts with axial mean temperature gradients is described by a second-order wave equation with variable coefficients.¹⁻⁴ Cummings¹ developed an approximate analytical solution of the wave equation for ducts with an arbitrary temperature gradients in the absence of mean flows. Munjal and Prasad⁵ and Peat⁶ have developed exact solutions for ducts with small temperature gradient, in the presence of mean flows. Also, Kapur *et al.*⁷ obtained numerical solutions for sound propagation in ducts with axial temperature gradients, in the

absence of mean flows, by integrating the wave equation using a Runge–Kutta method. The same approach was followed by Zinn *et al.*⁸⁻¹¹ who developed the impedance tube technique for high-temperature systems, in the presence of mean flows. Raspet *et al.*² obtained analytical solution for one-dimensional acoustic fields in the presence of a linear temperature gradient, in the absence of mean flow. Sujith *et al.*^{3,4} obtained exact analytical solutions for sound propagation in ducts with linear and exponential mean temperature profiles, and also for a temperature profile that corresponds to a constant convective heat transfer coefficient at the wall, in the absence of mean flow. The objective of this letter is to present exact analytical solutions for sound propagation in ducts with a quadratic mean temperature profile.

I. SOLUTION OF THE WAVE EQUATION

The one-dimensional wave equation for a constant area duct with a mean temperature gradient for a perfect, inviscid, and nonheat conducting gas, in the absence of mean flow is^{1,2}

$$\frac{\partial^2 p'}{dx^2} + \frac{1}{T} \frac{dT}{dx} \frac{\partial p'}{\partial x} - \frac{1}{\gamma R T} \frac{\partial^2 p'}{\partial t^2} = 0. \quad (1)$$

Assuming a solution of the form $p'(x, t) = P'(x)e^{i\omega t}$, Eq. (1) reduces to the following second-order ordinary differential equation for the complex amplitude $P'(x)$:

$$\frac{d^2 P'}{dx^2} + \frac{1}{T} \frac{dT}{dx} \frac{dP'}{dx} + \frac{\omega^2}{\gamma R T} P' = 0. \quad (2)$$

Sujith *et al.*⁴ have shown that Eq. (2) can be used for applications where the mean Mach number is smaller than 0.1.

In order to obtain an exact solution, Eq. (2) is rewritten with $\bar{T}(x)$ as the independent variable, yielding

$$\left(\frac{d\bar{T}}{dx}\right)^2 \frac{d^2 P'}{d\bar{T}^2} + \frac{1}{\bar{T}} \frac{d}{dx} \left(\bar{T} \frac{d\bar{T}}{dx}\right) \frac{dP'}{d\bar{T}} + \frac{\omega^2}{\gamma R} \frac{P'}{\bar{T}} = 0. \quad (3)$$

The above equation is solved for the case of duct with a quadratic mean temperature distribution that is given by the expression

$$\bar{T}(x) = ax^2 + bx + c. \quad (4)$$

To simplify Eq. (3) further, a new independent variable s is introduced:

$$s = -\frac{4a\bar{T}}{b^2 - 4ac}. \quad (5)$$

Noting that

$$\left(\frac{d\bar{T}}{dx}\right)^2 = 4a\bar{T} + (b^2 - 4ac), \quad (6)$$

$$\frac{d}{dx} \left[\bar{T} \frac{d\bar{T}}{dx} \right] = 6a\bar{T} + (b^2 - 4ac), \quad (7)$$

and transforming Eq. (3) to the s space yields

$$s(1-s) \frac{d^2 P'}{ds^2} + \left(1 - \frac{3}{2}s\right) \frac{dP'}{ds} - \frac{\omega^2}{4\gamma Ra} P' = 0. \quad (8)$$

Equation (8) is the hypergeometric differential equation,^{12,13} the standard form of which is given by

$$s(1-s) \frac{d^2 P'}{ds^2} + [\nu - (\alpha + \beta + 1)s] \frac{dP'}{ds} - \alpha\beta P' = 0, \quad (9)$$

where

$$\alpha = \frac{1}{4} + \frac{1}{2} \sqrt{\frac{1}{4} - \frac{\omega^2}{a\gamma R}}, \quad (10)$$

$$\beta = \frac{1}{4} - \frac{1}{2} \sqrt{\frac{1}{4} - \frac{\omega^2}{a\gamma R}}, \quad \text{and } \nu = 1.$$

This differential equation has three regular singular points $s = 0, 1$, and ∞ . The solution in the region $0 < s < 1$ is given by

$$P' = c_1 F(\alpha, \beta; 1; s) + c_2 F(\alpha, \beta; 1; s) \ln s$$

$$+ \sum_{n=1}^{\infty} \frac{(a)_n (b)_n}{(n!)^2} s^n [\psi(\alpha + n) - \psi(n) + \psi(\beta + n)$$

$$- \psi(\beta) - 2\psi(n+1) + 2\psi(1)], \quad (11)$$

where c_1 and c_2 are complex constants. ψ is the digamma function, which is given by

$$\psi(z) = d[\ln \Gamma(z)]/dz. \quad (12)$$

The solution for the acoustic pressure in the range of $1 < s < \infty$ is given by

$$P' = c_1 s^{-\alpha} F(\alpha, \alpha; \alpha - \beta + 1; 1/s)$$

$$+ c_2 s^{-\beta} F(\beta, \beta; \beta - \alpha + 1; 1/s). \quad (13)$$

II. CONCLUSIONS

An exact analytical solution describing the behavior of one-dimensional acoustic oscillations in ducts with a quadratic mean temperature profile is obtained. The analysis neglects mean flow effects and is therefore valid only for mean Mach numbers up to 0.1. This analysis should work well for pulse combustors where the temperature gradients are large, and the mean Mach numbers are small (less than 0.02), and also for engine exhaust systems with small mean Mach numbers. The analytical solutions presented herein can also be used as benchmarks for checking the results obtained from computer programs that are intended for applications for systems with axial mean temperature gradients and mean flow, in the limit as the mean Mach number becomes small.

¹A. Cummings, "Ducts with axial temperature gradients: An approximate solution for sound transmission and generation," *J. Sound Vib.* **51**, 55–67 (1977).

²R. Raspet, H. E. Bass, and J. Kordomenos, "Thermoacoustics of traveling waves: Theoretical analysis for an inviscid ideal gas," *J. Acoust. Soc. Am.* **94**, 2232–2239 (1993).

³R. I. Sujith, G. A. Waldherr, and B. T. Zinn, "Exact solution for one-dimensional acoustic fields in ducts with axial temperature gradient," Proceedings of the 32nd Aerospace Sciences Meeting, Reno, Nevada, 10–13 January, AIAA Paper No. 94-0359 (1994).

⁴R. I. Sujith, G. A. Waldherr, and B. T. Zinn, "Exact solution for one-dimensional acoustic fields in ducts with an axial temperature gradient," *J. Sound Vib.* **184**, 389–402 (1995).

⁵M. L. Munjal and M. G. Prasad, "On plane-wave propagation in a uniform pipe in the presence of a mean flow and a temperature gradient," *J. Acoust. Soc. Am.* **80**, 1501–1506 (1986).

⁶K. S. Peat, "The transfer matrix of a uniform duct with a linear temperature gradient," *J. Sound Vib.* **123**, 43–53 (1988).

⁷A. Kapur, A. Cummings, and P. Mungur, "Sound propagation in a combustion can with axial temperature and density gradients," *J. Sound Vib.* **25**, 129–138 (1972).

⁸M. Salikuddin and B. T. Zinn, "Adaptation of the impedance tube technique for the measurement of combustion process admittances," *J. Sound Vib.* **68**, 119–132 (1980).

⁹J. D. Baum, B. R. Daniel, and B. T. Zinn, "Determination of solid propellant admittances by the impedance tube technique," AIAA Paper No. 80-0281.

¹⁰B. T. Zinn and L. Narayanaswami, "Application of impedance tube technique in the measurement of driving provided by solid propellants during combustion instabilities," *Acta Astron.* **9**, 303–315 (1982).

¹¹L. M. Matta and B. T. Zinn, "Theoretical study of flow turning losses in the presence of temperature gradients," AIAA Paper No. 94-0099.

¹²M. Humi and W. Miller, *Second Course in Ordinary Differential Equations for Scientists and Engineers* (Springer-Verlag, New York, 1988).

¹³M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

Comments on “Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I and Part II” [J. Acoust. Soc. Am. 99, 1339–1348, 1349–1359 (1996)]

Jamal Assaad and Jean-Michel Rouvaen

Institut d'Electronique et de Microelectronique du Nord (IEMN), UMR CNRS 9929, OAE Department, Université de Valenciennes et du Hainaut Cambrésis, le Mont Houy, BP 311, 59304-Valenciennes Cedex, France

(Received 8 July 1996; accepted for publication 29 January 1997)

The aim of this Comment is to show that, when the normal velocity of the transducer is no longer constant, the algorithm that has been recently proposed by Wu *et al.* [J. Acoust. Soc. Am. 99, 1339–1348 for Part I and 1349–1359 for Part II (1996)] cannot eliminate aliasing nor remove the phase error in the angular spectrum. In fact, in recent articles, Wu *et al.* proposed and used a new numerical algorithm for eliminating the aliasing in the spectrum. Moreover, they have shown that the phase error can be easily removed. In their investigation they used a planar transducer and linear array. To show the effectiveness of their algorithm, the normal velocity of the planar transducer or of the elementary transducers (in the case of the array) has been supposed constant. © 1997 Acoustical Society of America. [S0001-4966(97)01506-3]

PACS numbers: 43.20.Rz, 43.20.Tb, 43.20.Bi [JEG]

INTRODUCTION

The purpose of this Comment is to demonstrate that the errors due to the sampling and the use of the discrete Fourier transform (DFT) cannot be eliminated when the normal velocity of the source is not a constant. In fact, the approach that Wu *et al.*^{1,2} proposed is based on the evaluation of two-dimensional acoustic fields. It has been shown that harmonic vibration can propagate at the surface of transducers.^{3,4} Thus in the general case the normal velocity of the transducer cannot be assumed constant.

In the case of an array, it is well known that the resultant pressure of an array mainly depends on the radiated pressure of an elementary transducer. Thus when we suppose that the normal velocity of the transducer or of the elementary transducer is constant (piston mode), the DSFT angular spectrum will give best results. This is clear by the inspection of Eq. (11) in the present paper. Moreover, Wu *et al.* say practically the same thing. In fact (see p. 1344 in Ref. 1), they have written, “the planar transducer with a width $2A$ and the normal velocity distribution $u(x)=1$ on its surface is a simple example showing the effectiveness of the DSFT: *The discrete form of the analytical Fourier transform of $v(x)$, Eq. (11), can be obtained by multiplying Eq. (9) with a sinc function and a phase shift term in Eq. (16).*” Thus to show the effectiveness of the DSFT in the case of piston mode the demonstration is obtained analytically and thus there is no need to make a lot of numerical calculations.

Wu's paper gives a good description of the angular spectrum approach (ASA). In addition, for the particular case of a piston mode, that paper gives an interesting algorithm. Moreover, Wu's paper gives an extensive review of the literature associated with the concept of ASA. Thus this description will not be recalled here. Finally, we will use the same notations as Ref. 1.

We think that using the finite element method together with ASA can be helpful especially in three-dimensional radiation problems. In fact, using the finite element method, we can compute the “true” normal velocity of the radiating structure in two- and three-dimensional radiation problems.³ Then applying the ASA, an approximated field can be computed at each point in the near-field and far-field domains.

I. SAMPLING APPROACHES

To use the ASA, the source (Fig. 1) must be sampled. As for signals there are a lot of methods to sample the normal velocity distribution $v(x)$ in the source plane $z=0$ (Fig. 1). The classical sampling of $v(x)$ can be mathematically described by

$$v^c(x) = \sum_{m_x=-N/2+1}^{N/2} v(m_x \Delta d) \delta(x - m_x \Delta d), \quad (1)$$

where δ is the Dirac function and Δd is the sampling interval (see Fig. 1). The sample and hold of $v(x)$ that has been proposed and used by Wu *et al.*¹ is given by

$$v^h(x) = \sum_{m_x=-N/2+1}^{N/2} v(m_x \Delta d) \delta(x - m_x \Delta d) * \text{rect}\left(\frac{x}{\Delta d}\right) \quad (2a)$$

with

$$\text{rect}\left(\frac{x}{\Delta d}\right) = 1 \quad \text{if } |x| \leq \frac{\Delta d}{2} \\ = 0 \quad \text{else}; \quad (2b)$$

where the asterisk denotes the convolution product. In the above two equations, the upper index (c or h) refers to the classical (c) and hold (h) sampling, respectively. This mathematical presentation is useful because it indicates the physi-

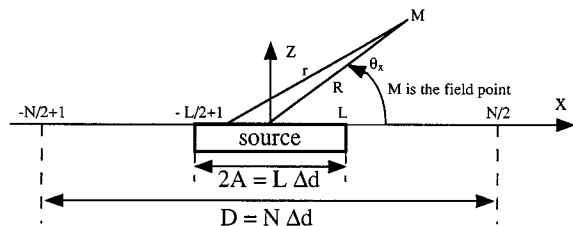


FIG. 1. Geometry and notation used for the calculation of the acoustic field.

cal interpretation of the used sampling. In fact, in Eq. (2) the value of $\nu(m_x \Delta d)$ is kept constant over the interval $[-\Delta d/2, \Delta d/2]$. Using the two previous expressions, it is simple to show that the discrete forms of the analytical Fourier transform (FT) of $\nu(x)$, are given by

$$V_D^c(m_x \Delta f) = \Delta d \sum_{i_x = -N/2+1}^{N/2} \nu(i_x \Delta d) \exp\left(\frac{-2j\pi m_x i_x}{N}\right) \quad (3)$$

and

$$V_D^h(m_x \Delta f) = V_D^c(m_x \Delta f) \text{sinc}\left(\frac{\pi m_x}{N}\right), \quad (4)$$

where $\Delta f = 1/(N\Delta d)$. Equation (3) is the DFT of the normal velocity and Eq. (4) gives the DSFT's (discrete sinc-Fourier transform) angular spectrum. Equation (4) has been obtained by Wu *et al.*¹ by using a somewhat more lengthy method. To remove the effect of the half-sample phase shift in the angular spectrum, Wu *et al.*^{1,2} proposed to multiply the previous equations by the $\exp(j2\pi m_x \phi/N)$ term. In this case Eq. (4) can be rewritten as

$$V_D^h(m_x \Delta f) = V_D^c(m_x \Delta f) \text{sinc}\left(\frac{\pi m_x}{N}\right) e^{j2\pi m_x \phi/N}, \quad (5)$$

where ϕ is equal to $1/2$ when L is even and equal to zero when L is odd ($2A = L\Delta d$, see Fig. 1). Using the above two expressions, the pressure field $p(i_x \Delta d, z)$ can be computed as

$$p_D^s(i_x \Delta d, z) = \frac{\Delta f \rho c}{\lambda} \sum_{m_x = -N/2+1}^{N/2} \frac{V_D^s(m_x \Delta f)}{U} \times \exp\left[2j\pi\left(\frac{i_x m_x}{N} + zU\right)\right], \quad (6a)$$

with

$$U = (1/\lambda)^2 - (m_x \Delta f)^2)^{0.5}, \quad (6b)$$

where the upper index s can be c , h , or e (e meaning "exact," see below), ρ is the fluid density, and c is the in-water sound velocity.

Let us assume that the normal velocity is equal to

$$\nu(x) = \cos(2\pi f_0 x) \text{rect}\left(\frac{x}{2A}\right) \quad \text{with } f_0 = k\Delta f \quad (k \in N), \quad (7)$$

where the $\text{rect}(x/2A)$ function is defined in Eq. (2b). The exact sampling of the analytical FT of the normal velocity is given by

$$V_D^e(m_x \Delta f) = \frac{L\Delta d}{2} \left[\text{sinc}\left(\frac{\pi L(m_x - k)}{N}\right) + \text{sinc}\left(\frac{\pi L(m_x + k)}{N}\right) \right]. \quad (8)$$

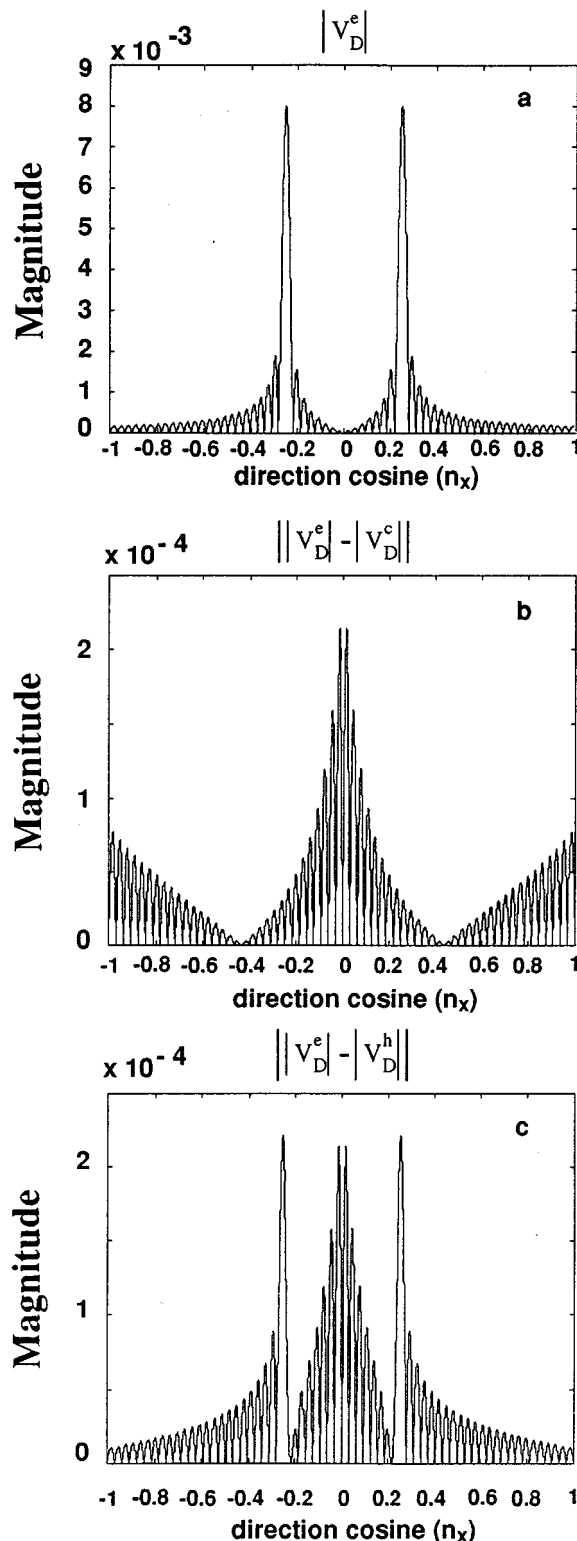


FIG. 2. The magnitude of the exact angular spectrum is given by (a). (b) and (c) display the difference between the exact and, respectively, the DFTs and DSFTs spectra. Vertical axes in (a), (b), and (c) are blown up by 10^3 , 10^4 , and 10^4 , respectively.

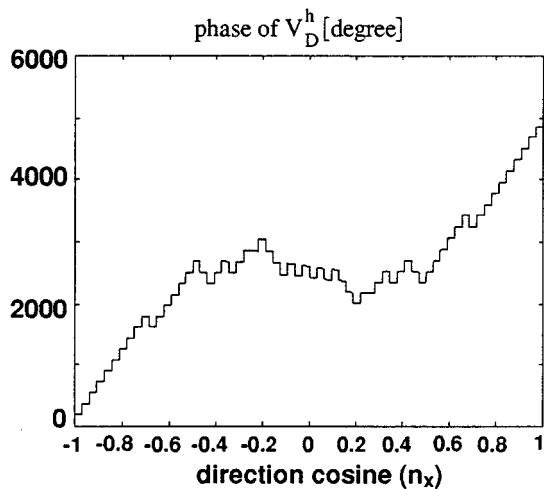


FIG. 3. The phase of the DSFT angular spectrum.

Using Eqs. (3), (5), (6), and (8) we can compute the pressure field in three different ways. When Eq. (8) is used together with Eq. (6) an expression very close to the analytical expression [Eq. (17) in Ref. 1] can be derived.

Using Eqs. (3) and (5), the DFT and DSFT can be easily computed as

$$V_D^c(m_x \Delta f) = \frac{L \Delta d}{2} \left[\left\{ \frac{\sin(\pi L(m_x - k))/N}{L \sin(\pi(m_x - k))/N} \right\} e^{-j2\pi(m_x - k)\phi/N} + \left\{ \frac{\sin(\pi L(m_x + k))/N}{L \sin(\pi(m_x + k))/N} \right\} e^{-j2\pi(m_x + k)\phi/N} \right] \quad (9)$$

and

$$V_D^h(m_x \Delta f) = V_D^c(m_x \Delta f) \operatorname{sinc}\left(\frac{\pi m_x}{N}\right) e^{j2\pi m_x \phi/N}. \quad (10)$$

If the normal velocity is constant (i.e., $k=0$), we immediately

$$V_D^e(m_x \Delta d) = V_D^h(m_x \Delta d). \quad (11)$$

It is clear that in this case the DSFT will give the best results and we can say that an “exact” angular spectrum is obtained. In the other case (i.e., k different from zero) the above equality is no more verified. Thus an exact angular spectrum cannot be obtained using the DSFT.

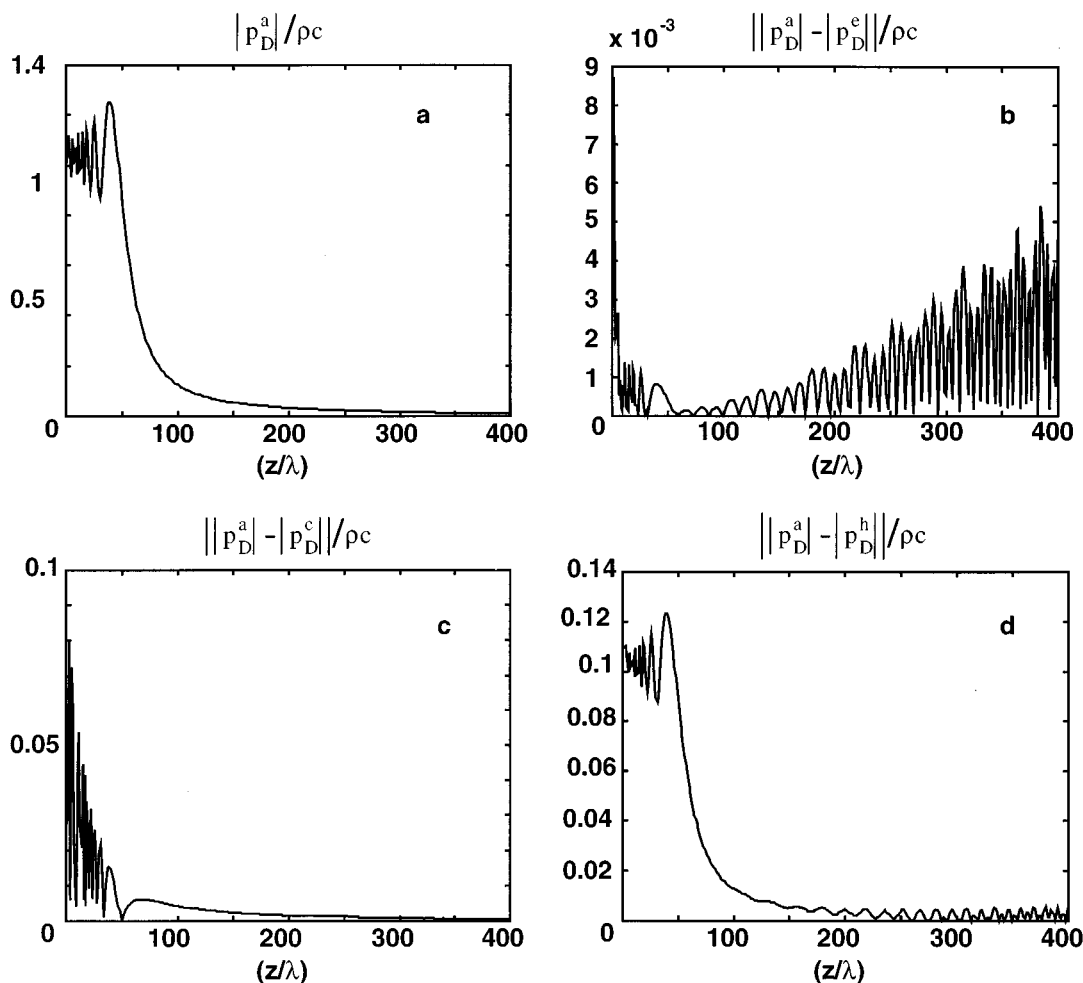


FIG. 4. (a) displays the exact analytical pressure magnitude. (b)–(d) give the difference between the analytical and the ASA’s pressure magnitudes. All figures are normalized by ρc . The vertical axis in (b) is blown up by 10^3 . Four hundred points have been used to draw each figure [i.e., $p(0, z)$ is computed for $z = \lambda$ to $z = 400\lambda$].

II. NUMERICAL RESULTS

The same parameters that have been used by Wu *et al.*¹ will be considered, i.e., $2A=32\lambda$ (λ being the wavelength), $\Delta d=\lambda/2$, $N=8192$, $L=64$, and $\lambda=0.5$ mm. But, the value of k is taken equal to $N/8$ (i.e., $1/f_0=\lambda_0=4\lambda$). Figure 2(a) shows the exact spectrum of the source [from Eq. (8)]. Figure 2(b) and (c) display the difference between the exact spectrum [Fig. 2(a)] and, respectively, the DFTs [from Eq. (9)] and DSFTs [Eq. (10)] angular spectra. It should be noted that for $n_x \approx \pm 0.25$ ($\theta_x \approx 75.5^\circ$ for $n_x \approx +0.25$ and $\theta_z = \pi/2 - \theta_x$) the error is important when the DSFT angular spectrum is used. θ_x and θ_z give the orientation of the wave vector \mathbf{k} with respect to the x and z coordinate axes (Fig. 1). In addition, the far-field pressure exhibits a main lobe in the directions $\beta = \pm \arcsin(\lambda f_0)$ approximately equal to $\pm 14.5^\circ$ [see Eq. (B6) in Ref. 4]. Thus $\beta = \theta_z$.

Figure 3 shows the unwrapped phase spectrum of the DSFT [from Eq. (5)], which is clearly not nullified. Thus shifting the phase in the angular spectrum by the $\exp(j\pi n_x/N)$ term in Eq. (4), when the normal velocity is not a constant, cannot remove the half sample length phase shift error.

Using the analytical expression [Eq. (14) from Ref. 1], we calculate the field on the z axis (i.e., $x=0$) and display it in Fig. 4(a). Figure 4(b)–(d) give the difference between the field presented in Fig. 4(a) and the pressure field, obtained by using, respectively, the exact [Eq. (8)], DFT [Eq. (9)], and

DSFT [Eq. (10)] angular spectrum. It is interesting to note that when the DSFT angular spectrum is used, the difference is more important.

III. CONCLUSION

By reading Wu's papers, we may think that the conclusions drawn on the study of simple piston mode may be extended to any type of transducer. In the present paper, the normal velocity has been supposed harmonic. Such a situation can physically exist (see Ref. 4). We think that some clarifications have been given. Finally, the limitation of DSFT has been shown and the DFT spectrum still gives the best results in some cases.

¹P. Wu, R. Kazys, and T. Stepinski, "Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I. Errors due to the discrete Fourier transform and discretization," *J. Acoust. Soc. Am.* **99**, 1339–1348 (1996).

²P. Wu, R. Kazys, and T. Stepinski, "Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part II. Characteristics as a function of angular range," *J. Acoust. Soc. Am.* **99**, 1349–1359 (1996).

³J. Assaad, J.-N. Decarpigny, C. Bruneel, R. Bossut, and B. Hamonic, "Application of the finite element method to two-dimensional radiation problems," *J. Acoust. Soc. Am.* **94**, 562–573 (1993).

⁴J. Assaad, C. Bruneel, J.-N. Decarpigny, and B. Nongaillar, "Electromechanical coupling coefficients and far-field radiation patterns of lithium niobate bars (Y-cut) used in high-frequency acoustical imaging and non-destructive testing," *J. Acoust. Soc. Am.* **94**, 2969–2978 (1993).

Response to “Comments on ‘Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I and Part II’” [J. Acoust. Soc. Am. 101, 3800 (1997)]

Ping Wu, Rymantas Kazys, and Tadeusz Stepinski

Uppsala University, Department of Technology, Circuits and Systems, Box 534, S-751 21 Uppsala, Sweden

(Received 30 January 1997; accepted for publication 13 February 1997)

The authors respond to the comments of Assaad and Rouvaen. © 1997 Acoustical Society of America. [S0001-4966(97)01606-8]

PACS numbers: 43.20.Rz, 43.20.Tb, 43.20.Bi [JEG]

Assaad and Rouvaen discuss some issues related to our paper.¹ They claim namely that when the normal velocity of the transducer surface is variable (we assumed a constant velocity in our paper), the algorithm referred to in our paper as the discrete sinc Fourier transform (DSFT) can neither eliminate aliasing nor remove the phase error in the angular spectrum. Below we try to clarify these issues.

Concerning the removal of the phase error, we are afraid that a small error in the formulation presented by Assaad and Rouvaen has led to a wrong conclusion. As for the elimination of the aliasing, the conclusions that can be drawn depend on the interpretation of the aliasing error and the requirements concerning the computation algorithm to attain a satisfactory accuracy when computing the angular spectrum.

We will try to prove the above statements using the same notation as in the Comments and in our paper.¹ For the source with a nonconstant normal velocity considered by Assaad and Rouvaen [see Eq. (7) in the Comments], the discrete form of the normal velocity is written as follows:

$$v(i_x \Delta d) = \begin{cases} \cos(2\pi f_0(i_x - \phi)\Delta d), \\ -L/2 + 1/2 + \phi \leq i_x \leq L/2 - 1/2 + \phi, \\ 0, \text{ otherwise,} \end{cases} \quad (1)$$

or

$$v(i_x \Delta d) = \sum_{n=-L/2+1}^{L/2} \cos(2\pi f_0(i_x - \phi)\Delta d) \delta[(i_x - n)\Delta d] \otimes \text{rect}\left(\frac{i_x \Delta d}{\Delta d}\right) \quad (1')$$

[cf. Eq. (2a) in the Comments], where $f_0 = k\Delta f$ ($k \in N$), \otimes denotes spatial convolution, L is integer, and ϕ is 1/2 when L is even and zero when L is odd. In other words, if sampled in an even number, the source is shifted by a half-sample length, $\Delta d/2$ (Δd is the sampling interval) and if sampled in an odd number, it is not shifted.

The DFT of the discretized normal velocity in Eq. (1), i.e., the DFT-based angular spectrum, takes the following form:

$$\begin{aligned} V_D^c(m_x \Delta f) &= \Delta d \sum_{i_x=-L/2+1}^{L/2} \cos\left(\frac{2\pi k}{N}(i_x - \phi)\right) \exp\left[\frac{-j2\pi m_x i_x}{N}\right] \\ &= \frac{L\Delta d}{2} \left\{ \frac{\sin[(\pi/N)(m_x - k)L]}{L \sin[(\pi/N)(m_x - k)]} \right. \\ &\quad \left. + \frac{\sin[(\pi/N)(m_x + k)L]}{L \sin[(\pi/N)(m_x + k)]} \right\} \exp\left(\frac{-j\pi m_x}{N} \phi\right), \quad (2) \end{aligned}$$

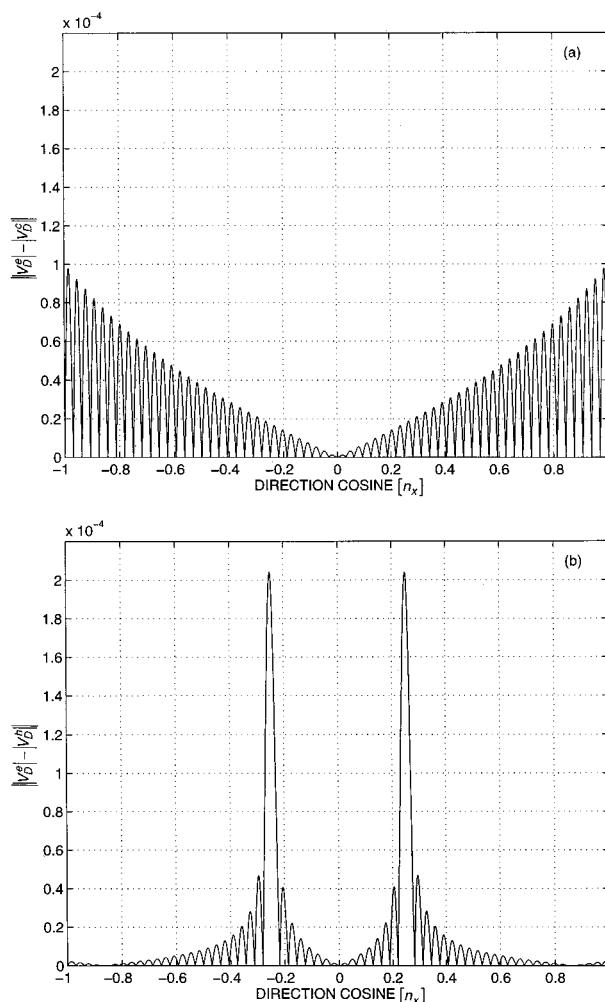


FIG. 1. The correct results of (a) $\|V_D^c(m_x \Delta f) - |V_D(m_x \Delta f)|\|$ and (b) $\|V_D^c(m_x \Delta f) - |V_D^h(m_x \Delta f)|\|$.

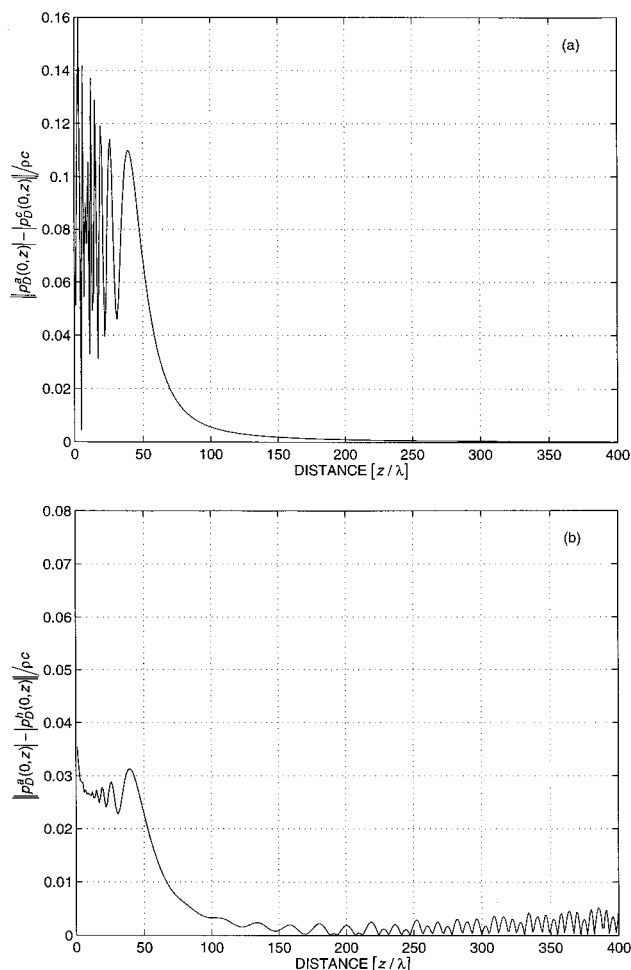


FIG. 2. The correct results of (a) $\|p_D^a(0,z) - |p_D^c(0,z)|\|/pc$ and (b) $\|p_D^a(0,z) - |p_D^h(0,z)|\|/pc$.

which is different from Eq. (9) in the Comments. If Eq. (2) is multiplied by $\exp(j\pi m_x \phi/N)$, the phase error with reference to the analytical form [Eq. (8) in the Comments] will be corrected. Multiplying Eq. (2) with $\exp(j\pi m_x \phi/N) \text{sinc}(\pi m_x/N)$, we obtain the DSFT-based angular spectrum with correction of phase error, $V_D^h(m_x \Delta f)$.

The corrected results obtained using the same parameters as in the Comments, i.e., $2A=32\lambda$, $\Delta d=\lambda/2$, $N=8192$, $\lambda=0.5$ mm, and $k=N/8$ ($1/f_0=\lambda_0=4\lambda$), are presented in Fig. 1 where $\|V_D^a(m_x \Delta f) - |V_D^c(m_x \Delta f)|\|$ and $\|V_D^e(m_x \Delta f) - |V_D^h(m_x \Delta f)|\|$ are plotted as functions of direction cosine. Using the corrected $V_D^c(m_x \Delta f)$ and $V_D^h(m_x \Delta f)$, we have calculated $p_D^c(m_x \Delta f)$ and $p_D^h(m_x \Delta f)$. The results expressed as a difference from the analytical solution $p_D^a(m_x \Delta f)$, which is given in Ref. 1, are shown in Fig. 2.

The results here in Figs. 1 and 2 are different from those in Figs. 2 and 4 in the Comments and thus lead us to the different conclusions. A comparison of (a) and (b) in Fig. 1 shows that the DFT-based angular spectrum has larger error than the DSFT-based one except in the neighborhood of $n_x=\pm 0.25$. A comparison of (a) and (b) in Fig. 2 demonstrates that much larger error appears in the DFT-based field than in the DSFT-based one except that, in the DSFT-based field, some small ripples appear beyond $z=100\lambda$. The Comments concerned only one case where $f_0=1/4\lambda$ and the fields calculated were on the axis. The Comments by Assaad and Rouvaen encouraged us to clarify this matter in an extended study of the DSFT applied to the sources of the above studied type.

¹P. Wu, R. Kazys, and T. Stepinski, "Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I. Errors due to the discrete Fourier transform and discretization," and "Part II. Characteristics as a function of angular range," J. Acoust. Soc. Am. **99**, 1339–1359 (1996).

Effect of monaural and binaural altered auditory feedback on stuttering frequency

Andrew Stuart, Joseph Kalinowski, and Michael P. Rastatter

Department of Communication Sciences and Disorders, East Carolina University, Greenville,
North Carolina 27858-4353

(Received 29 August 1996; accepted for publication 28 February 1997)

The effect of monaural and binaural alterations in auditory feedback on stuttering frequency was investigated. Eleven participants who stutter read aloud under nonaltered auditory feedback (NAF) and monaural and binaural conditions of frequency altered feedback [(FAF), one-quarter octave shift upward] and delayed auditory feedback [(DAF), 50-ms delay] at a normal speech rate. Relative to the NAF condition, reductions in stuttering frequency of approximately 60%–75% were found with the altered auditory feedback conditions. *Post hoc* single-df comparisons revealed a reduction in stuttering frequency with altered auditory feedback versus NAF ($p < 0.0001$), a greater reduction in stuttering frequency for binaural compared to monaural altered auditory feedback ($p = 0.028$), and nonsignificant differences in stuttering frequencies for right versus left monaural conditions ($p = 0.54$) and DAF versus FAF ($p = 0.70$). © 1997 Acoustical Society of America. [S0001-4966(97)05006-6]

PACS numbers: 43.70.Dn, 43.66.Pn [AL]

INTRODUCTION

Alterations in auditory feedback have been known for decades to reduce stuttering frequency for those who stutter. The introduction of auditory masking or masked auditory feedback (MAF), delayed auditory feedback (DAF), and frequency-altered feedback (FAF) have been demonstrated to reduce stuttering frequency relative to nonaltered auditory feedback (NAF) (see Bloodstein, 1995 for a review). DAF and FAF have been shown to be more effective in reducing stuttering frequency than MAF (Howell *et al.*, 1987; Kalinowski *et al.*, 1993). Whereas Howell *et al.* (1987) reported FAF to be more effective than DAF in reducing stuttering frequency, others have reported DAF and FAF to be equally effective (Kalinowski *et al.*, 1993; MacLeod *et al.*, 1995). Although both DAF and FAF are effective in reducing stuttering frequency, the effect of each on speech production is different: Individuals who stutter increase their voice level under DAF but decrease it under FAF (Howell, 1990). Further, the proportion of dysfluent prolongations for stuttering speakers increases and decreases under DAF and FAF, respectively (Howell *et al.*, 1988).

A wearable prosthetic device employing altered auditory feedback has been suggested as an adjunct or alternative to current stuttering therapy (e.g., Armson *et al.*, 1995; Kalinowski *et al.*, 1995). That is, a therapeutic approach using a prosthetic device may be more beneficial for those individuals who have difficulty producing natural sounding speech following therapy (e.g., Franken *et al.*, 1992; Kalinowski *et al.*, 1994; Martin *et al.*, 1984; Metz *et al.*, 1990; Runyan and Adams, 1979; Runyan *et al.*, 1990) and/or those having difficulty maintaining “motoric strategies” (e.g., Boberg, 1981). Any prosthetic device should be “acoustically invisible” (Armson *et al.*, 1995; Stuart *et al.*, 1996). In other words, the device should reproduce the high fidelity of unaided listening and auditory self monitoring while at the same time delivering optimal altered feedback. Armson *et al.*

(1995) and Stuart *et al.* (1996) have operationally defined optimal altered feedback as those parameters which maximize stuttering reduction while at the same time are minimally distorted (i.e., sounds closest to NAF). Several reports have been directed to exploring the minimal FAF shifts and DAF delays that are the most fluency enhancing (Hargrave *et al.*, 1994; Kalinowski *et al.*, 1996; Stuart *et al.*, 1996). To date, a 50-ms delay in auditory feedback and frequency shifts in altered auditory feedback as little as plus/minus one-quarter of an octave can be considered optimal (Kalinowski *et al.*, 1996; Stuart *et al.*, 1996).

One query that remains is whether monaural altered auditory feedback is as effective in reducing stuttering frequency as binaural altered auditory feedback. In other words, would one prosthetic device be as effective as two? Traditionally, binaural altered feedback has been employed and there is a paucity of studies exploring the efficacy of monaural altered auditory feedback. Barr and Carmel (1969) first reported the effect of monaural high-frequency narrow-band MAF on stuttering frequency relative to NAF. Ten participants who stuttered displayed a reduction in stuttering frequency under the two conditions of narrow-band MAF. Unfortunately, the authors failed to conduct inferential statistical analyses and as such the significance of stuttering reduction is not known. Further, they did not include a binaural MAF condition. Yari (1976) investigated the effect of binaural and monaural broadband white noise at a 90-dB sensation level with MAF with a NAF control condition. Six adults who stuttered displayed a reduction in stuttering in all MAF conditions relative to the NAF condition. The binaural MAF condition produced a significant reduction in stuttering frequency compared to NAF. Although there was a reduction in stuttering frequency observed in both monaural conditions, compared to the NAF condition, the difference did not attain statistical significance.

To date there have been no studies that have reported the

effects of other forms of monaural altered auditory feedback on stuttering frequency. As noted above, DAF and FAF have proven to be more efficacious than MAF in reducing stuttering frequency in those that stutter. The purpose of this study was, therefore, to investigate the efficacy of yet unexplored monaural DAF and FAF relative to binaural DAF, FAF, and NAF. Specifically, an investigation was undertaken of the effect of monaural and binaural frequency alterations of a plus one-quarter octave shift and a 50-ms delay in auditory feedback on stuttering frequency.

I. METHODS

A. Participants

Participants were eight adults and three adolescents who stutter (9 males and 2 females, Mean=20.9 yr, Standard error=2.1). All participants had a history of therapy although none were enrolled at the time of testing. Participants were recruited at the State University of New York at Geneseo prior to beginning an intensive therapy program. Participants did not present with any other speech or language problems. All participants had normal bilateral hearing sensitivity, defined as having thresholds of 25 dB HL or better at octave frequencies from 250 to 8000 Hz (American National Standards Institute, 1996), and normal bilateral middle-ear function (American Speech-Language-Hearing Association, 1990).

B. Apparatus

Participants, while seated in a double-walled sound-treated audiometric suite, spoke into a microphone 15 cm from their mouths (JVC model M-510) held with a boom on a stand with an approximate orientation of 330° azimuth and -30° altitude. The microphone output was fed to an audio mixer (Studiomaster model session mix) and routed to a digital signal processor (Yamaha model DSP-1) and amplifier (Sansui model AU-55900) before being returned to the participants' ears through insert earphones (EAR Tone model 3A). Speech input was routed through the processor unaltered for the NAF condition. In the FAF conditions, the entire speech signal was shifted in frequency up one-quarter octave. In the DAF conditions the speech signal was delayed 50 ms. The amplifier gain for speech input was set such that the output to the insert earphones was consistent with auditory self-monitoring during normal conversation. That is, the output of the insert earphones was calibrated such that a speech input of 75 dB SPL (consistent with an average conversation level at 15 cm from a speaker's mouth) had an output of 85 dB SPL in a standard 2-cm³ HA-2 coupler (consistent with an average real ear conversation SPLs of speech outputs from normal speakers).¹ Participants' speech samples were video recorded with a camera (Quasar model X8CCD) and video stereo cassette recorder (JVC model BR-6400U).

C. Procedures

Participants read seven different junior-high level passages of slightly more than 300 syllables with similar theme

and syntactic complexity. Participants were instructed to read aloud at a normal rate (i.e., at their "usual" or "normal" reading rate) for all seven experimental conditions: NAF and monaural (both right and left ears) and binaural DAF and FAF. Participants were instructed not to use any strategies or techniques to control or reduce dysfluencies. The NAF condition served as a baseline and preceded the six conditions of altered auditory feedback. The presentation order of the six conditions of altered auditory feedback was counterbalanced. During the altered feedback conditions participants received either binaural or monaural feedback. For the monaural conditions one insert earphone was removed from the nontest ear. In an attempt to minimize carryover stuttering reduction effects from previous auditory feedback conditions, participants were asked to produce approximately two to three minute monologues between reading passages. All participants were tested in one session which lasted approximately 45 min.

Stuttering episodes were calculated from the first 300 syllables of participants' videotape recorded passages by a trained research assistant who was unaware of the purpose of the study. Stuttering was defined as part-word repetitions, part-word prolongations, and/or inaudible postural fixations. Stuttering frequency was recalculated by the same research assistant for 10% of the speech samples chosen at random. Intrajudge syllable by syllable agreement was 0.80, as indexed by Cohen's kappa (Cohen, 1960). Cohen's kappa values above 0.75 represent excellent agreement beyond chance (Fleiss, 1981). A certified speech-language pathologist, who is highly experienced in the area of fluency and fluency disorders, independently determined stuttering frequency for 10% of the speech samples chosen at random. Interjudge syllable by syllable agreement, was 0.79 as indexed by Cohen's kappa.

II. RESULTS

Stuttering frequencies for each participant as a function of auditory feedback condition are presented in Table I. Means and standard errors for stuttering frequency (i.e., number of stuttering episodes/300 syllables) as a function of auditory feedback condition are shown in Fig. 1. Compared to NAF, reductions in stuttering frequency of approximately 60% and 75% occurred for the monaural and binaural conditions, respectively.

A multivariate one-factor analysis of variance (ANOVA) with repeated measures was performed to investigate the effect of auditory feedback condition on stuttering frequency. The ANOVA revealed a statistically significant main effect of auditory feedback condition [$F(6,66) = 14.05$, $p < 0.0001$, Huynh-Feldt $p < 0.0001$, $\omega^2 = 0.50$]. A number of orthogonal single-df comparisons were undertaken to assess the differences between treatment means of interest (Keppel, 1991; Keppel and Zedeck, 1989). There was a significant reduction in stuttering frequency for all forms of altered auditory feedback relative to NAF [$F(1,60) = 78.67$, $p < 0.0001$, Huynh-Feldt $p < 0.0001$, $\omega^2 = 0.78$]. There were also statistically significant fewer stuttering episodes with binaural compared to monaural altered auditory feedback [$F(1,60) = 5.07$, $p = 0.028$, Huynh-

TABLE I. Stuttering frequencies (i.e., number of stuttering episodes/300 syllables) as a function of auditory feedback for each participant. NAF=nonaltered auditory feedback, FAF=frequency altered feedback, DAF=delayed auditory feedback, and L, R, B=left, right, and binaural, respectively.

| Participant | NAF | DAF-R | DAF-L | DAF-B | FAF-R | FAF-L | FAF-B |
|-------------|-----|-------|-------|-------|-------|-------|-------|
| 1 | 19 | 5 | 13 | 1 | 7 | 17 | 4 |
| 2 | 16 | 12 | 10 | 12 | 5 | 3 | 1 |
| 3 | 44 | 11 | 7 | 4 | 5 | 6 | 6 |
| 4 | 26 | 1 | 4 | 7 | 2 | 6 | 3 |
| 5 | 13 | 7 | 6 | 8 | 6 | 11 | 6 |
| 6 | 15 | 9 | 9 | 5 | 2 | 8 | 5 |
| 7 | 22 | 6 | 6 | 4 | 17 | 7 | 8 |
| 8 | 71 | 57 | 66 | 33 | 50 | 52 | 37 |
| 9 | 43 | 2 | 4 | 4 | 8 | 8 | 3 |
| 10 | 16 | 3 | 3 | 1 | 6 | 1 | 2 |
| 11 | 12 | 1 | 0 | 1 | 0 | 0 | 1 |

Felt $p=0.043$, $\hat{\omega}^2=0.16$]. Nonsignificant differences in stuttering frequencies were found for right versus left monaural conditions [$F(1,60)=0.39$, $p=0.54$, Huynh-Felt $p=0.45$, $\hat{\omega}^2=0.00$] and DAF versus FAF [$F(1,60)=0.70$, $p=0.70$, Huynh-Felt $p=0.59$, $\hat{\omega}^2=0.00$].

III. CONCLUSIONS AND DISCUSSION

The findings of the present study demonstrate the following: Altered forms of auditory feedback (i.e., monaural or binaural DAF and FAF) significantly reduced stuttering frequency relative to NAF. Binaural altered auditory feedback was more effective than monaural altered auditory feedback in reducing stuttering frequency. DAF and FAF were equally efficacious in reducing stuttering frequency. Finally, there was no difference between the right or left monaural altered auditory feedback conditions.

These outcomes are in agreement with previous studies that have reported reductions in stuttering frequency for monaural altered auditory feedback relative to NAF, albeit a significantly smaller reduction compared to binaural altered

auditory feedback (Barr and Carmel, 1969; Yari, 1976). These results also support the previous finding that there are no significant differences between right and left monaural altered auditory feedback (Barr and Carmel, 1969). Consistent with the findings of Kalinowski *et al.* (1993) and MacLeod *et al.* (1995), DAF and FAF were proven to be equally efficacious in stuttering reduction among individuals who stutter.

The finding of an advantage in binaural over monaural altered auditory feedback in reducing stuttering frequency is perplexing. Although there is evidence to support a relationship between the sensation level of altered auditory feedback and stuttering reduction (Adams and Hutchinson, 1974; Conture, 1974; Maraist and Hutton, 1957; Shane, 1957), it seems unlikely that this may be offered as a plausible explanation for the advantage that binaural altered auditory feedback has over monaural altered auditory feedback. That is, it seems unlikely that the three to six decibel enhancement in an auditory signal due to binaural summation of loudness (e.g., Hirsh, 1948) could account for the significant reduction in stuttering frequency observed with various binaural versus the same monaural forms of altered auditory feedback. We speculate that the binaural advantage observed with DAF, MAF, and FAF may be the coherence of the auditory feedback signals to that part of the central nervous system underlying the amelioration of stuttering. In other words, during binaural altered auditory feedback two coherent fluency enhancing auditory signals drive that part of the central nervous system responsible for stuttering reduction. During monaural altered auditory feedback two auditory feedback signals compete: an altered auditory feedback signal that induces stuttering reduction and a normal NAF signal that does not induce stuttering reduction. One could suggest that the competition or incoherence of the NAF signal reduces the effectiveness of the fluency enhancing altered auditory feedback and is therefore responsible for the lesser reductions in stuttering frequency observed in monaural versus binaural DAF, FAF, and MAF conditions. In other words, a binaural coherent altered auditory feedback signal offers a better substrate for the underlying component in the central nervous system responsible for stuttering reduction.

The ability of altered auditory feedback to immediately reduce stuttering suggests that the use of a prosthetic device

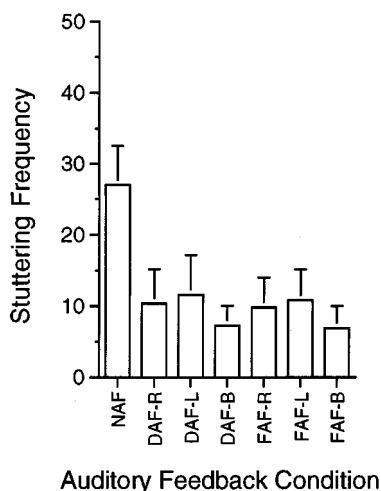


FIG. 1. Mean stuttering frequencies (i.e., number of stuttering episodes/300 syllables) as a function of auditory feedback ($n=11$). Error bars represent plus one standard error. NAF=nonaltered auditory feedback, FAF=frequency altered feedback, DAF=delayed auditory feedback, and L, R, B=left, right, and binaural, respectively.

employing these effects may be an effective means for achieving stuttering reduction outside the therapeutic environment. These findings suggest that employing a device(s) with binaural altered auditory feedback would be the most efficacious in reducing stuttering frequency. It is recognized that the above speculation must be tempered as further data is necessary to assess the naturalness of speech produced under conditions of altered auditory feedback, the effects of prolonged exposure to alter auditory feedback, the acceptability of the altered auditory feedback to the individual who stutters, and whether the effects of altered auditory feedback generalize to the population of situations an individual who stutters experiences in daily life. There is preliminary data to suggest that for some speech produced under altered auditory feedback is natural sounding (Ingham *et al.*, 1997; White *et al.*, 1995) and stuttering frequency is reduced in difficult speaking situations outside the laboratory environment (Armson *et al.*, in press; Zimmerman *et al.*, in press). We are presently involved in additional investigations aimed at assessing the feasibility of employing altered auditory feedback outside the therapeutic environment.

ACKNOWLEDGMENTS

The assistance of Dr. Kathleen Jones, State University of New York at Geneseo, Geneseo, NY in participant recruitment is greatly appreciated. In addition, we wish to thank Dr. D. P. Phillips for valuable discussion during the preparation of the manuscript and Dr. Peter Howell and an anonymous reviewer for their insightful comments during the review process.

¹See Kalinowski *et al.* (1993) for a detailed description of calibration procedures.

- Adams, M. R., and Hutchinson, J. (1974). "The effects of three levels of auditory masking on selected vocal characteristics and the frequency of dysfluency of adult stutterers," *J. Speech Hear. Res.* **17**, 682–688.
- ANSI (1996). ANSI S3.6-1996, "American National Standard Specifications for Audiometers" (American National Standards Institute, New York).
- American-Speech-Language-Hearing Association (1990). "Guidelines for screening for hearing impairments and middle ear disorders," *Asha* **32** (Suppl. 2), 17–24.
- Armson, J., Kalinowski, J., Foote, S., Witt, C., and Stuart, A. (in press). "Effect of frequency altered feedback and audience size on stuttering," *Eur. J. Dis. Commun.*
- Armson, J., Kalinowski, J., and Stuart, A. (1995). "A model of stuttering remediation: multiple factors underlying fluency enhancement," in *Stuttering: Proceedings from the First World Congress on Fluency Disorders*, edited by C. W. Starkweather and H. F. M. Peters (Nijmegen U. P., Nijmegen, The Netherlands), pp. 296–300.
- Barr, D. F., and Carmel, N. R. (1969). "Stuttering inhibition with high frequency narrow-band masking noise," *J. Aud. Res.* **9**, 40–44.
- Bloodstein, O. (1995). *A Handbook on Stuttering* (The National Easter Seal Society, Chicago), 5th ed., pp. 327–357.
- Boberg, E. (1981). *Maintenance of Fluency* (Elsevier North-Holland, New York).
- Cohen, J. (1960). "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.* **20**, 37–46.
- Conture, E. G. (1974). "Some effects of noise on the speaking behavior of stutterers," *J. Speech Hear. Res.* **17**, 714–723.
- Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (Wiley, New York), 2nd ed., pp. 212–236.
- Franken, M. C., Boves, L., Peters, H. F. M., and Webster, R. (1992). "Per-

- ceptual evaluation of the speech before and after fluency shaping therapy," *J. Fluency Dis.* **17**, 223–242.
- Hargrave, S., Kalinowski, J., Stuart, A., Armson, J., and Jones, K. (1994). "Effect of frequency altered feedback on stutterers' fluency at two speech rates," *J. Speech Hear. Res.* **37**, 1313–1319.
- Hirsh, I. J. (1948). "Binaural summation: A century of investigation," *Psychol. Bull.* **45**, 193–206.
- Howell, P. (1990). "Changes in voice level caused by several forms of altered auditory feedback in fluent speakers and stutterers," *Lang. Speech* **33**, 325–338.
- Howell, P., El-Yaniv, N., and Powell, D. J. (1987). "Factors affecting fluency in stutterers," in *Speech Motor Dynamics in Stuttering*, edited by H. F. M. Peters and W. Hulstijn (Springer-Verlag, New York), pp. 361–369.
- Howell, P., Wingfield, T., and Johnson, M. (1988). "Characteristics of the speech of stutterers during normal and altered auditory feedback," in *Proceedings Speech 88*, edited by W. A. Ainsworth and J. N. Holmes (Institute of Acoustics, Edinburgh), Vol. 3, pp. 1069–1076.
- Ingham, R. J., Moglia, R. A., Frank, P., Costello-Ingham, J., and Cordes, A. (1997). "Experimental investigation of the effects of frequency-altered feedback on the speech of adults who stutter," *J. Speech Hear. Res.* **40**, 361–372.
- Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A., and Gracco, V. (1993). "Effects of alterations in auditory feedback and speech rate on stuttering frequency," *Lang. Speech* **36**, 1–16.
- Kalinowski, J., Armson, J., Stuart, A., Hargrave, S., Sark, S., and MacLeod, J. (1995). "Effect of alterations in auditory feedback on stuttering frequency during fast and normal speech rates," in *Stuttering: Proceedings from the First World Congress on Fluency Disorders*, edited by C. W. Starkweather and H. F. M. Peters (Nijmegen U. P., Nijmegen, The Netherlands), pp. 51–55.
- Kalinowski, J., Noble, S., Armson, J., and Stuart, A. (1994). "Naturalness ratings of the pretreatment and posttreatment speech of adults with mild and severe stuttering," *Am. J. Speech-Lang. Path.* **3**, 65–70.
- Kalinowski, J., Stuart, A., Sark, S., and Armson, J. (1996). "Stuttering amelioration at various feedback delays and speech rates," *Eur. J. Dis. Commun.* **31**, 259–269.
- Keppel, G. (1991). *Design and Analysis: A Researcher's Handbook* (Prentice-Hall, Englewood Cliffs, NJ), 3rd ed., pp. 344–366.
- Keppel, G., and Zedeck, S. (1989). *Data Analysis for Research Design* (Freeman, New York), pp. 261–291.
- Maraist, J. A., and Hutton, C. (1957). "Effects of auditory masking upon speech of stutterers," *J. Speech Hear. Dis.* **22**, 385–389.
- Martin, R. R., Haroldson, S. K., and Triden, K. A. (1984). "Stuttering and speech naturalness," *J. Speech Hear. Dis.* **49**, 53–58.
- Metz, D. E., Schiavetti, N., and Sacco, P. R. (1990). "Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers," *J. Speech Hear. Dis.* **55**, 516–525.
- MacLeod, J., Kalinowski, J., Stuart, A., and Armson, J. (1995). "Effect of single and combined altered auditory feedback on stuttering frequency at two speech rates," *J. Commun. Dis.* **28**, 217–228.
- Runyan, C., and Adams, M. R. (1979). "Unsophisticated judges' perceptual evaluations of the speech of 'successfully treated' stutterers," *J. Fluency Dis.* **4**, 29–48.
- Runyan, C. M., Bell, J. N., and Prosek, R. A. (1990). "Speech naturalness ratings of treated stutterers," *J. Speech Hear. Dis.* **55**, 434–438.
- Shane, M. L. S. (1957). "Effect on stuttering of alteration in auditory feedback," in *Stuttering in Children and Adults: Thirty Years of Research at the University of Iowa*, edited by W. Johnson and R. R. Luetenegger (University of Minnesota, Minneapolis, MN), pp. 286–297.
- Stuart, A., Kalinowski, J., Armson, J., Stenstrom, R., and Jones, K. (1996). "Fluency effect of frequency alterations of plus/minus one-half and one-quarter octave shifts in auditory feedback of people who stutter," *J. Speech Hear. Res.* **39**, 396–401.
- White, C., Kalinowski, J., and Armson, J. (1995). "Clinicians' naturalness ratings of stutterers' speech under auditory feedback conditions," *Asha* **37**(10), 107.
- Yari, E. (1976). "Effects of binaural and monaural noise on stuttering," *J. Aud. Res.* **16**, 114–119.
- Zimmerman, S., Kalinowski, J., Stuart, A., and Rastatter, M. (in press). "Effect of altered auditory feedback on people who stutter during scripted telephone conversations," *J. Speech Hear. Res.* (to be published).

Erratum: “Vibrational impact of high-speed trains. I. Effect of track dynamics” [J. Acoust. Soc. Am. 100, 3121–3134 (1996)]

Victor V. Krylov

*Centre for Research into the Built Environment, Nottingham Trent University, Burton Street, Nottingham
NG1 4BU, United Kingdom*

(Received 14 January 1997; accepted for publication 20 February 1997)

[S0001-4966(97)01306-4]

PACS numbers: 43.40.At, 43.40.Kd, 43.10.Vx [PJR]

In the above-mentioned paper, the transition from the calculated dynamic values of track deflection curves $w(\nu t)$ to the corresponding load forces $P(t)$ applied from sleepers to the ground at very high train speeds has been made using the earlier derived¹ quasi-static relationship $P(t) = T[2w(\nu t)/w_{\max}^{\text{st}}](d/x_0)$ —formula (5) of the paper. This led to a wrong conclusion that the approaching of the train speed to the critical track wave velocity will be associated with a decreased level of generated ground vibrations. In fact, the result is opposite: the level of generated vibrations will *increase* when train speeds approach the track wave critical velocity.

To provide the correct result, formula (5) should be re-written in a more general form valid for both quasi-static and dynamic regimes:

$$P(t) = T[2w(\nu t)/w_{\max}^{\text{st}}](d/x_0^{\text{st}}). \quad (5)$$

Here, w_{\max}^{st} and x_0^{st} are, respectively, the maximum value of $w(\nu t)$ and the track deflection length, both being calculated in the quasi-static approximation. Derivation of (5) goes straightforward if we take into account that $P(t) = \alpha w(\nu t)\Delta d$, where Δd is the sleeper width and α is the constant of the Winkler foundation. Exclusion of α and Δd from this expression may be done by integrating Eq. (1) of the paper taking into consideration a quasi-static regime

(i.e., for $\partial^2 w / \partial t^2 = 0$) over the distance along the track x . The integration, which takes into account that load forces are transmitted to the ground through sleepers only, results in the formula $\alpha w_{\max}^{\text{st}} \Delta d N_{\text{eff}}^{\text{st}} = T$, where the quantity $N_{\text{eff}}^{\text{st}} = \pi/2\beta d = x_0^{\text{st}}/2d$ determined as a solution of Eq. (6) of the paper is the effective number of sleepers equalizing the quasi-static axle load T . Combined with the expression for $P(t)$, this gives the above written more general formula (5).

Since, for dynamic values of $w(\nu t)$, the factor $\delta = (1 - \nu^2/c_{\min}^2)^{1/2}$ is present in the denominator of the corrected formula (5) [and of the associated formula (10)], the forces $P(t)$ applied from sleepers to the ground (Fig. 3) will increase as the train speed approaches the critical track wave velocity. Therefore, in Figs. 6 and 11 of the paper and in the corresponding conclusion one will have the increase in levels of generated ground vibrations for train speeds approaching the critical track wave velocity, instead of the earlier predicted decrease. The corrected Figs. 3, 6, and 11 will be reproduced in detail in a forthcoming publication. Other formulas, figures, and conclusions of the paper remain unchanged.

¹V. V. Krylov, “Generation of ground vibrations by superfast trains,” *Appl. Acoust.* **44**, 149–164 (1995).

TECHNICAL NOTES AND RESEARCH BRIEFS

Paul B. Ostergaard

10 Glenwood Way, West Caldwell, New Jersey 07006

Editor's Note: Original contributions to the Technical Notes and Research Briefs section are always welcome. Manuscripts should be double-spaced, and ordinarily not longer than about 1500 words. There are no publication charges, and consequently, no free reprints; however, reprints may be purchased at the usual prices.

Advanced-degree dissertations in acoustics

Editor's note: Abstracts of Doctoral and Master's theses will be welcomed at all times. Please note that they must be double spaced, limited to 200 words, must include the appropriate PACS classification numbers, and formatted as shown below (don't make the editor retype them, please!). The address for obtaining a copy of the thesis is helpful. Please submit two copies.

Acoustic finite-element calculations in the time domain

[43.20.Fn, 43.20.Px, 43.20.Tb]—Morten Skaarup Jensen, *Department of Acoustic Technology, Technical University of Denmark, February 1997 (Ph.D.)*. Transient sound fields can be predicted by solving a problem consisting of a second-order hyperbolic partial-differential equation with appropriate initial and boundary conditions. This thesis is concerned with the numerical solution of such problems. The finite-element method is used for the spatial discretization, and various methods including the Newmark method are used for the time discretization. The most serious error in time-domain finite-element calculations reveals itself as false dispersion: An emitted pulse will gradually be distorted, because different frequency com-

ponents of the signal are predicted to travel with different speed. This dispersion error can be reduced by using smaller elements and a finer time discretization, but this is very costly. Several methods aimed directly at reducing the dispersion error have been examined, but although good suppression of the error was observed in special cases, no method of general validity has been found. However, if the only way of reducing the error is to use finer spatial and temporal discretization, it becomes essential to reduce the computational costs without sacrificing accuracy. Accordingly, a number of methods of reducing computational costs are suggested, analyzed, and tested. Some of these methods seem to be very promising. Theoretical considerations on determining the optimum element size and time step are presented. It is also concluded that traditional direct-frequency domain solutions require more computational work than frequency domain solutions obtained by fast Fourier transformation of time-domain solutions.

Thesis advisor: Finn Jacobsen.

Copies of this thesis may be obtained from the Department of Acoustic Technology, Technical University of Denmark, Building 352, DK-2800 Lyngby, Denmark.