CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 114, No. 1 July 2003

ACOUSTICAL NEWS—USA 1

USA Meetings Calendar 2

ACOUSTICAL STANDARDS NEWS 7

Standards Meetings Calendar 7

REPORTS OF RELATED MEETINGS 17

ABSTRACTS FROM ACOUSTICS RESEARCH LETTERS ONLINE 19

BOOK REVIEWS 21

REVIEWS OF ACOUSTICAL PATENTS 23

LETTERS TO THE EDITOR

Scattering of acoustic waves from immersed transversely isotropic cylinders (L)	Farhang Honarvar, Ying Fan, Anthony N. Sinclair	45
Time-domain pressure response of arrays with periodic excitation (L)	Jeffrey A. Ketterling, Frederic L. Lizzi	48
Reciprocity and the active control of distributed acoustic sources (L)	Stephen J. Elliott, Lewis Rees	52
Adaptive plasticity of loudness induced by chronic attenuation and enhancement of the acoustic background (L)	C. Formby, L. P. Sherlock, S. L. Gold	55
Second derivative analysis of consonant-vowel transition waveforms (L)	K. H. Norian	59
The effect of trabecular material properties on the frequency dependence of backscatter from cancellous bone (L)	Keith A. Wear	62
Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations (L)	Douglas P. Nowacek, Brandon M. Casper, Randall S. Wells, Stephanie M. Nowacek, David A. Mann	66
Estimation of the Nakagami parameter from log-compressed ultrasonic backscattered envelopes (L)	P. M. Shankar	70

GENERAL LINEAR ACOUSTICS [20]

Acoustic wave propagation in double porosity media	Xavier Olny, Claude Boutin	73
Bifurcation of the Biot slow wave in a porous medium	Inna Edelman	90

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 1, JULY 2003

CONTENTS—*Continued from preceding page*

Empirical model of the acoustic impedance of a circular orifice in grazing mean flow — Seong-Hyun Lee, Jeong-Guon Ih — 98

NONLINEAR ACOUSTICS [25]

The shock formation distance in a bounded sound beam of finite amplitude — Chao Tao, Jian Ma, Zhemin Zhu, Gonghuan Du, Zihong Ping — 114

UNDERWATER SOUND [30]

Ray stability in weakly range-dependent sound channels — F. J. Beron-Vera, M. G. Brown — 123

Horizontal array beamforming in an azimuthally anisotropic internal wave field — Steven Finette, Roger Oba — 131

Robust time reversal focusing in the ocean — Seongil Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, T. Akal — 145

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

A guided wave technique for the characterization of highly attenuative viscoelastic materials — F. Simonetti, P. Cawley — 158

Theory of the Greenspan viscometer — Keith A. Gillis, James B. Mehl, Michael R. Moldover — 166

TRANSDUCTION [38]

Acoustic characteristics of an electrodynamic planar digital loudspeaker — Kenji Furihata, Atsushi Hayama, David K. Asano, Takesaburo Yanagisawa — 174

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Edge waves in poroelastic plate under plane stress conditions — P. Malla Reddy, M. Tajuddin — 185

An approximate Green's function for a locally excited fluid-loaded thin elastic plate — Daniel T. DiPerna, David Feit — 194

NOISE: ITS EFFECTS AND CONTROL [50]

Transmission loss predictions for dissipative silencers of arbitrary cross section in the presence of mean flow — Ray Kirby — 200

Improving robustness of active noise control in ducts — Jing Yuan — 210

Noise duration for a single overflight — Rufin Makarewicz, Hanna Wojciechowska — 218

ACOUSTIC SIGNAL PROCESSING [60]

Synthetic aperture imaging using sources with finite aperture: Deconvolution of the spatial impulse response — Fredrik Lingvall, Tomas Olofsson, Tadeusz Stepinski — 225

Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix — Claire Prada, Jean-Louis Thomas — 235

PHYSIOLOGICAL ACOUSTICS [64]

Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves — Christopher A. Shera — 244

Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears — Michael P. Gorga, Stephen T. Neely, Darcia M. Dierking, Patricia A. Dorn, Brenda M. Hoover, Denis F. Fitzpatrick — 263

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 1, JULY 2003

CONTENTS—*Continued from preceding page*

Suppression tuning in noise-exposed rabbits	MacKenzie A. Howard, Barden B. Stagner, Paul K. Foster, Brenda L. Lonsbury-Martin, Glen K. Martin	279
Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects	Van Summers, Egbert de Boer, Alfred L. Nuttall	294
Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey	Mitchell Steinschneider, Yonatan I. Fishman, Joseph C. Arezzo	307

PSYCHOLOGICAL ACOUSTICS [66]

The effects of a high-frequency suppressor on tuning curves and derived basilar-membrane response functions	Ifat Yasin, Christopher J. Plack	322
An account of monaural phase sensitivity	Robert P. Carlyon, Shihab Shamma	333
Asymmetry of masking between complex tones and noise: Partial loudness	Hedwig Gockel, Brian C. J. Moore, Roy D. Patterson	349
Examination of a linear model in an informational masking study	Zhongzhou Tang, Virginia M. Richards	361
Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity	Nathaniel I. Durlach, Christine R. Mason, Barbara G. Shinn-Cunningham, Tanya L. Arbogast, H. Steven Colburn, Gerald Kidd, Jr.	368
Effectiveness of narrow-band versus tonal off-frequency maskers	Sophie Savel, Sid P. Bacon	380
The effects of the amplitude distribution of equal energy exposures on noise-induced hearing loss: The kurtosis metric	Roger P. Hamernik, Wei Qiu, Bob Davis	386
Intensity discrimination and increment detection in cochlear-implant users	Magdalena Wojtczak, Gail S. Donaldson, Neal F. Viemeister	396
Perceived naturalness of spectrally distorted speech and music	Brian C. J. Moore, Chin-Tuan Tan	408
Precedence-effect thresholds for a population of untrained listeners as a function of stimulus intensity and interclick interval	Kourosh Saberi, Joseph V. Antonio	420
Vertical-plane sound localization probed with ripple-spectrum noise	Ewan A. Macpherson, John C. Middlebrooks	430
Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers	Michael K. Qin, Andrew J. Oxenham	446

SPEECH PRODUCTION [70]

Modeling vocal fold motion with a hydrodynamic semicontinuum model	M. Drew LaMar, Yingyong Qi, Jack Xin	455
Production benefits of childhood overhearing	Leah M. Knightly, Sun-Ah Jun, Janet S. Oh, Terry Kit-fong Au	465

SPEECH PERCEPTION [71]

Dichotic speech recognition in noise using reduced spectral cues	Philipos C. Loizou, Arunvijay Mani, Michael F. Dorman	475
The effect of speech presentation level on measurement of auditory acclimatization to amplified speech	Kevin J. Munro, Mark E. Lutman	484

MUSIC AND MUSICAL INSTRUMENTS [75]

Delayed models for simplified musical instruments	Ana Barjau, Vincent Gibiat	496
The design of bells with harmonic overtones	Neil McLachlan, Behzad Keramati Nigjeh, Anton Hasell	505

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 1, JULY 2003

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

Acoustic communication in two freshwater gobies: Ambient noise and short-range propagation in shallow streams	M. Lugli, M. L. Fine	512
A diminished rate of "physiological decay" at noise offset contributes to age-related changes in temporal acuity in the CBA mouse model of presbycusis	James R. Ison, Paul Allen	522
Whole-lung resonance in a bottlenose dolphin (*Tursiops truncatus*) and white whale (*Delphinapterus leucas*)	James J. Finneran	529
Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear wave speeds in muscles	Jean-Luc Gennisson, Stefan Catheline, Sana Chaffaï, Mathias Fink	536
Applications of laser-based ultrasonics to the characterization of the internal structure of teeth	D. W. Blodgett	542

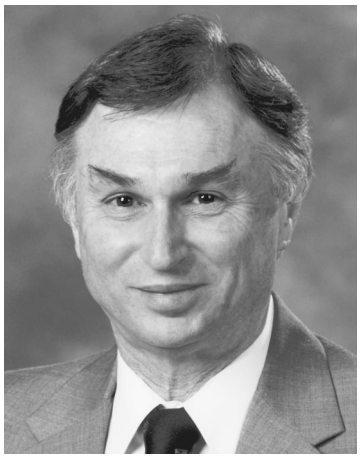
CUMULATIVE AUTHOR INDEX 550

ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America



Hans P. Gottlieb—For contributions to the mathematical treatment of vibrational phenomena.

Jules S. Jaffe—For contributions to acoustic imaging.

Oswald Leroy—For contributions to acousto-optics and interfacial waves.

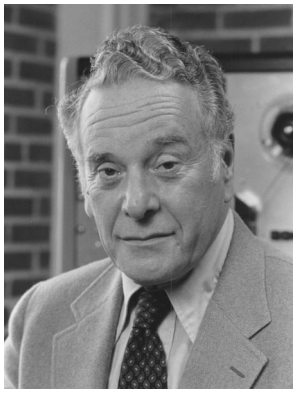
Raymond Lim—For contributions to the theory of acoustic scattering from objects buried in the seafloor.

Zoi-Heleni Michalopoulou—For contributions to signal processing and mathematical analysis of ocean data.

Mario A. Svirsky—For contributions to understanding speech perception in children and adults with cochlear implants.

 0001-4966/2003/114(1)/1/5/$19.00

Ira Hirsh receives lifetime achievement award



Ira J. Hirsh, Ph.D., has received the Peter H. Raven Lifetime Achievement Award from the Academy of Science of St. Louis. The award was presented on 2 April 2003.

Hirsh has been integral to the programs of Central Institute for the Deaf (CID) and Washington University since 1951. He has held numerous positions at both institutions during the past 52 years, including director of CID and dean of the faculty of arts and sciences at Washington University. Currently, he is the Edward Mallinckrodt Distinguished Professor Emeritus of Psychology and Audiology at Washington University, a professor in the CID-based speech and hearing department, and director of research emeritus, director emeritus and a member of the board of managers at CID.

Hirsh has published the textbook, *The Measurement of Hearing*, and more than 100 scholarly articles covering topics such as the measurement of hearing, sound conduction in the ear, cochlear mechanics, masking, auditory localization and perception, psychoacoustic behavior in animals and humans, speech communication, audiometry, psychophysics of hearing and ecological acoustics.

In 2002, the Acoustical Society of America (ASA) hosted a half-day symposium honoring Hirsh's research contributions. Other honors include a Gold Medal from ASA, a Whetnall Medal from the Royal Society of Medicine, Honors of the American Speech-Language-Hearing Association, a Biennial Award from ASA, a Distinguished Alumnus award from the State University of New York at Albany, and an Alumnus Merit Award from Northwestern University.

Hirsh is a member of the National Academy of Sciences and a fellow of the Acoustical Society of America, the American Psychological Association, the American Psychological Society, the American Speech and Hearing Association, the American Association for the Advancement of Science and the Society of Experimental Psychologists. He also has served as chair of the National Research Council's Commission on Behavioral and Social Sciences and Education (1982–1987) and on the American National Standards Institute and International Commission on Acoustics. Hirsh holds master's and doctoral degrees in psychology from Harvard University, a master's degree in speech from Northwestern University and a bachelor's degree in English and math from State University of New York State at Albany.

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.

2003

27–30 July	1st Conference on Acoustic Communication by Animals, University of Maryland, College Park, MD [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org/communication.html].
4–6 Aug.	46th meeting of the Acoustic Emission Working Group (AEWG), Portland, OR [Richard Nordstrom; E-mail: nordsr@cecs.pdx.edu; WWW: www.cecs.pdx.edu/~nordsrAEWG46/aewg46.html].
5–8 Oct.	IEEE International Ultrasonics Symposium, Honolulu, HI [W. D. O'Brien, Jr., Bioacoustics Research Lab., Univ. of Illinois, Urbana, IL 61801-2991; Fax: 217-244-0105; WWW: www.ieee-uffc.org].
10–14 Nov.	146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
3–7 Aug.	8th International Conference of Music Perception and Cognition, Evanston, IL [School of Music, Northwestern Univ., Evanston, IL 60201; WWW: www.icmpc.org/conferences.html].
15–19 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

Revision List

New Associates

Antonelli, George A., Intel Corp., 5200 NE Elam Young Pkwy., RA1-234, Hillsboro, OR 97124
Belanger, Zackery B., Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607
Beltman, Willem M., 18298 Upper Midhill Dr., West Linn, OR 97068
Brock, Daniel K., 4700 SW 17th St., Topeka, KS 66604-2454
Brunt, Christopher S., P.O. Box 325, Topanga, CA 90290
Chapelon, Jean-Yves, INSERM, U556, 151 Sours Albert Thomas, Lyon 69424, France
Chu, Wai C., 1336 Tanaka Dr., San Jose, CA 95131
Conroy, James P., Ensco, Inc., 5400 Port Royal Rd., Springfield, VA 22151-2388
Crawford, John J., Auburn Engineered Audio and Acoustics, 557 Old Stage Rd., Auburn, AL 36830
Day, Thomas W., Musictech College, 19 East Exchange St., Saint Paul, MN 55101
Diamond, Steven S., AGI, Inc., 1010 Tyinn St., #24, Eugene, OR 97402
Dietz, Jerrold E. G., National Security Dept., Johns Hopkins Applied Physics Lab., 11100 Johns Hopkins Rd., MS 8-220, Laurel, MD 20723-6099
Dong, David W., Veneklasen Associates, 1711 Sixteenth St., Santa Monica, CA 90404
Duncan, Michael E., Planning Systems Inc., 115 Christian Ln., Slidell, LA 70458
Farquharson, James G., Faszer Farquharson & Associates, Ltd., Suite 304, 6051 St., SW, Calgary, AB T2P 3S9, Canada
Feist, Jeffrey P., Polaris Industries, Inc., Victory Motorcycles R&D, 805 Seminole Ave., Osceola, WI 54020
Foley, Dennis L., ASP #149184 H-1, B-24, P.O. Box 324 Winslow/Apache, St. Johns, AZ 85936-3240
Huelskemper, Michael, Siemens AG, ICM MP P TI2, Frankenstrasse 2, Bocholt NRW 46395, Germany
Kinney, Jack W., 5579 Kingsbury Rd., Fairfield, OH 45014
Laws, Robert A., 39 Park Rd., Watford WD117 4QW, England
Leary, Del M., Porous Materials, National Center for Physical Acoustics, 1 Coliseum Dr., University, MS 38677
Lewis, George K., 5 Bayberry Lane, Andover, MA 01810
Lin, Feng, 110 Corrina Blvd., Apt. 360, Waukesha, WI 53186
Mao, Chi-Min, 23 Daffodil Rd., Rocky Point, NY 11778
Marty, Pierre N., CSM Materialteknik, Gelbjutaregatan, Linkoping, Ostergotlan 58183, Sweden
Mascolo, Antonio, Via Marina di Puolo 1, Massa Lubrense, Naples 80061, Italy
McGough, Robert J., Electrical and Computer Engineering, Michigan State Univ., 2120 Engineering Bldg., East Lansing, MI 48824
Meek, James A., 1523 East Anderson Dr., Palatine, IL 60074
Moulton, Dave L., Thale Acoustics, Waverley Industrial Park, Haisham Dr., Harrow HA1 4TR, United Kingdom
Nielsen, Aaron P., EVI Technology, 7138 Columbia Gateway Dr., Columbia, MD 21046
Nita, Bogdan G., Physics, Univ. of Houston, 617 Science and Research Bldg. 1, Houston, TX 77204
Oxley, Brylie, 7427 Overton Ave., Apt. 5, Raytown, MS 64133
Panzer, Joerg W., 19 Windermere, Huntingdon PE29 6UE, United Kingdom
Paul, Aoki M., Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304
Paulauskis, John A., HBE Corp., 11330 Olive St., St. Louis, MO 63141
Pedigo, Alan B., Hart Freeland Roberts, Inc., 7101 Executive Center Dr., Suite 300, Brentwood, TN 37027
Poroy, Oguz, Biomedical Eng., 1136 Seamans Center, Iowa City, IA 52242
Radentz, Michael G., JMR Holdings LLC, 47 Willmore Rd., St. Louis, MO 63109
Randolph, Patricia, Div. of Otolaryngology, Howard Univ. Hospital, 2041 Georgia Ave., NW, Suite 4B27, Washington, DC 20060
Ridings, Jason J., Russ Berger Design Group, 4006 Beltline Rd., Suite 160, Addison, TX 75001
Ross, Annie, Faculte d'Ingenierie, Univ. de Moncton, Moncton, NB E1A 3E9, Canada
Ruffa, Francisco, Gabriela Mistral 4240, Buenos Aires 1419, Argentina
Sandoe, Iain D., Sandoe Acoustics, Ltd., 94A Greatbridge Rd., Romsey, Hampshire SO51 8FH, United Kingdom
Schiff, Martin T., Cerami & Associates, Inc., 317 Madison Ave., #220, New York, NY 10017-5201
Scofield, Glenn A., Clear Gravy Productions, P.O. Box 270, Frederick, CO 80530-0270
Sharp, David B., Environmental and Mechanical Eng., Open Univ., DEME, Faculty of Eng., Walton Hall, Milton Keynes MK7 6AA, United Kingdom
Shaw, Richard M., Campbell-Shaw, Inc., 6461 Penfield Ave., Woodland Hills, CA 91367-2721
Simons, Richard D., Latin Percussion, 160 Belmont Ave., Garfield, NJ 07026
Sinclair, Alexandra M., 41 Alameda Padre Serra, Santa Barbara, CA 93103
Sweitzer, Glenn E., 1601 Superior Dr., #9, Auburn, IN 46706
Teeple, Ryan J., The Drucker Company, 200 Shady Ln., Philipsburg, PA 16866
Valk, Henry S., 3032 St. Helena Dr., Tucker, GA 30084-2227
Walsh, Kenneth M., K+M Engineering Ltd., 51 Bayberry Ln., Middletown, RI 02842
White, Jeffrey R., Soneticom, 1045 South John Rodes Blvd., West Melbourne, FL 32904
Willems, Stefan M. J., Philips, Inteleuvenlaan 74, Leuven B-3001, Belgium
Zanelli, Claudio L., Onda Corp., 592 Weddell Dr., Suite 7, Sunnyvale, CA 94089
Zhang, Xiangling, GM's Electro-Motive Div., Dept. 835K, 9301 55th St., La Grange, IL 60525-8051

New Students

Armstrong, Peter R., MIT, 77 Massachusetts Ave., Rm. 5-418, Cambridge, MA 02139
Balani, Aditya G., 10 Vairo Blvd., 231A, State College, PA 16803
Bernstein, Joshua G., Research Lab. of Electronics, Massachusetts Inst. of Technology, 77 Massachusetts Ave., 36-765, Cambridge, MA 02139
Bonati, Joshua T., 311 Cathedral St., Apt. 5C, Baltimore, MD 21201
Caronti, Alessandro, Electronic Engineering, Univ. Rome Tre, Via della Vasca Navale 84, Roma 00146, Italy
Cheon, Sang Yee, 2717 Pamoa Rd., Honolulu, HI 96822
Chia, Natasha J., Engineering Science, Univ. of Oxford, Parks Rd., Oxford OX1 3PJ, United Kingdom
Chupp, Aaron A., 432 South 2nd St., West Dundee, IL 60118
Cox, Keith L., 204 Maplewood Circle, Griffin, GA 30224
Danforth, Amanda, Pennsylvania State Univ., P.O. Box 30, State College, PA 16804
Daniels, Michelle, P.O. Box 15872, Stanford, CA 94309
Doty, Benjamin J., Graduate Program in Acoustics, Pennsylvania State Univ., P.O. Box 30, State College, PA 16804
Evans, Bronwen G., 2 Cavendish Close, St. John's Ave., Putney, London SW15 2AL, United Kingdom
Freyaldenhoven, Melinda C., 3500 Sutherland Ave., Apt. J208, Knoxville, TN 37919
Fukaya, Teruhiko, Linguistics Dept., Univ. of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693
Garson, Charles J., 1140 East Meadowmere, Springfield, MO 65807
Gauthier, Philippe-Aubert, C.P. 4199, Lac-Beauport, PQ G0A 2C0, Canada
Ghoshal, Goutam, Engineering Mechanics, Univ. of Nebraska—Lincoln, W317.4 NH Eng. Mechanics, Lincoln, NE 68588
Hess, Wolfgang G., Inst. of Communication Acoustics, Ruhr-Univ., Universitaetsstr. 150, Bochum NRW 44780, Germany
Hurley, Shawn, 3245 Dutch Ridge Rd., Staitsville, OH 43766
Imbrie, Annika K., Massachusetts Inst. of Technology, 77 Massachusetts Ave., Cambridge, MA 02139
Ku, Emery M., Swarthmore College, 500 College Ave., Swarthmore, PA 19081
Lee, Hotaik, 1400 Martin St., #1092, State College, PA 16803

Lee, Robert, 1316 Crosswinds Court, #2, Lawrence, KS 66046
Maul, Kristen K., Speech and Hearing Sciences, City Univ. of New York, 365 Fifth Ave., New York, NY 10016
May-Bowers, Carie L., 1613 Bayo Vista Ave., San Pablo, CA 94806
Mechling III, Hiram K., 1005 Marengo, New Orleans, LA 70115
Mehta, Tejas, Electrical Engineering, Univ. of Southern California, Los Angeles, CA 90089
Niehus, Rebecca, Linguistics, Univ. of Michigan, 4080 Frieze Bldg., 105 South State St., Ann Arbor, MI 48109
Novascone, Stephen R., 3022 Jackie Court, Idaho Falls, ID 83402
Nykaza, Edward T., 268 Toftrees Ave., #123, State College, PA 16803
Ozer, Mehmet B., Mechanical and Industrial Eng. Dept., Univ. of Illinois at Chicago, M/C 251, 2039 ERF, 842 West Taylor St., Chicago, IL 60613
Parham, Douglas F., School of Audiology and Speech-Language Pathology, The Univ. of Memphis, 807 Jefferson Ave., Memphis, TN 38105
Phatak, Sandeep, 512 East Stoughton St., Apt. 311, Champaign, IL 61820
Philipp, Norman H., 1320 Ash St., Eudora, KS 66025
Renfro, Jason M., 1390 Athens Dr., Palm Bay, FL 32907
Ritchie, Peter J. Ritchie, 1130 Magnolia Dr., Cleveland, OH 44106
Rogers, Deanna S., Speech and Hearing Sciences, Indiana Univ., 200 North Jordan Ave., Bloomington, IN 47405
Snow, Evan, Vtech, 212 Thomas, Blacksburg, VA 24006
Sokolow, Adam, 34 Avonmore Way, Penfield, NY 14526
Spahr, Erik J., MWC Box 1488, 1701 College Ave., Fredericksburg, VA 22401
Sullivan, Sarah C., Psychology Dept., Univ. of Texas at Austin, 1 University Station A8000, Austin, TX 78712-0187
Tan, Lee, Computer Engineering Dept., Price of Songkhla Univ., Hat Yai Songkhla 90112, Thailand
Tavernier, Daniel P., Bethel College, Physics Dept., 3900 Bethel Dr., #2681, Arden Hills, MN 55112
Theis, Kevin R., Zoology, Michigan State Univ., Natural Science Bldg., Rm. 203, East Lansing, MI 48824-1115
Thorn, Aisha, CCNR, Univ. of Sussex, School of Cognitive Sciences, Brighton BN1 9QG, United Kingdom
van Waterschoot, Toon J. M., ESAT-SCD/SISTA, Kasteelpark Arenberg 10, Heverlee B-3001, Belgium
Velazquez, Marvin, 206B White Course D, University Park, PA 16802-6821
Wagner, James Wagner, 1761 Forest Dr., Medina, OH 44256
Wang, Hai, Aerospace and Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02215
Webster, Joanna D., 810 Douglas Dam Rd., New Market, TN 37820
Wilbur, Jed C., Aerospace and Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02215
Zhao, Xialiang, 6020 Pure Sky Place, Clarksville, MD 21029
Zhi, Ni, 650 Merrick, Apt. B-12, Detroit, MI 48202

New Electronic Associates

Bamberger, Andreas T., Inst. of Physics, Univ. Freiburg, Hermann-Herderstr. 3, Freiburg 79104, Germany
Bozeman, John K., Private Bag 302-149, North Harbour, Auckland 1330, New Zealand
Capleton, Brian, The Coach House, Blackheath Way, West Malvern, Worcestershire WR14 4DR, United Kingdom
Culbertson, Deborah S., Communication Sciences and Disorders, East Carolina Univ., Corner of Chalres and Greenville Blvds., Greenville, NC 27858
Daigle, Marie A., Speech and Language Pathology, Maplesheade Elementary School, 175 Mapleshade Ave., East Longmeadow, MA 01028
Fisher, Karl A., Lawrence Livermore National Lab., NTED, 7000 ''E'' Ave., Livermore, CA 94551
Giron, Franck, Advanced Software Lab., Sony Corporate Laboratories Europe, Hedelfingerstrasse 61, Stuttgart 70327, Germany
Grawunder, Sven, Georg-Cantor Str. 23, Halle 06108, Germany
James, Christopher J., 19 rue Marc Sangnier, Toulouse 31300, France
Makepeace, Shawn L., 5 Carson Dr., Belleville, IL 62223
Masta, Robert I., Interstellar Research, 2247 Prairie St., Ann Arbor, MI 48105
Militello, Carmelo, Physics Dept., Univ. of La Laguna, Astrofisico F. Sanchez S/N, La Laguna 38200, Spain
Price, Marshall, 2280 SW 26th Ln., Miami, FL 33133-2330
Sveshnikov, Boris V., 1 Timiryazev St., P.O. Box 72, Nizhny, Novgorod 603022, Russia
Vornbrock, Ted J., 903 Erie Ave., Takoma Park, MD 20912

Associates to Members

Y. Arieh, M. R., Bailey, B. T. Cox, H. A. Drotleff, C. Fox, S. A. Fulop, H. E. Gockel, G. A. Haverstick, C. N. Himmel, R. C. W. Ip, B. S. Johnson, T. E. Johnson, B. L. Masiello, T. J. Matula, T. Montlick, B. C. Mueller, G. Naylor, D. J. Oldham, C. J. Pollock, P. Ratilal, M. T. Shaw, J. N. Soker, R. M. Su, M. C. M. Wright, M. Xu, Z. Zhang

Students to Associates

B. L. Adams, J. A. Corey, J. J. DiGiovanni, M. A. Epstein, R. J. Fedewa, B. L. Fedor, M. J. Greaves, M. C. Kelly, S. Kim, O. D. Kripfgans, R. A. Moscoso-Bullon, S. Nissen, D. E. Scott, K. D. Wallace, Y. Zheng

Associate to Student

P. T. Calamia

Member to Electronic Associate

H. F. Voigt

Associates to Electronic Associates

A. Alencar, J. S. Allen, C. G. Fox, T. Haresign, T. M. Huber, D. Kapolka, S. Mehta, T. L. Nelson, S. J. Smith

Student to Electronic Associate

A. N. Zagrai

Reinstated

J. Lee—*Member*

Resigned

R. Efron, F. Iversen—*Fellows*

J. A. Doutt, M. Hollins, S. Lees, P. E. Lyregaard, D. N. MacLennan, L. R. Moss, J. D. Revell, A. F. Russell, J. Utman—*Members*

E. S. Bishop, D. Delisi, H. J. Eom, L. R. Kahn, H. L. Zhou—*Associates*

F. Fiates, S. Kurra, C. Tong, M. D. Tuel-Benckendorf—*Students*

A. Krasnic—*Electronic Associate*

Deceased

P. Leehey, R. H. Mellen, E. B. Yeager—*Fellows*

D. R. Heebner, H. A. Leeper—*Members*

H. A. Hair—*Associate*

Dropped

Martin B. Barmatz, Evan F. Berkman, Walter L. Clearwaters, John R. Franks, Jean-Luis Guyader, Richard N. Hamme, Lucio Maestrello, Otohiko Nomoto, David R. Perrott, Gerard J. Quentin, Wolfgang H. Sachse, Robert Y. Ting, Michael F. Werby—*Fellows*

Pedro L. M. Albernaz, John A. Arthur, Carl W. Asp, Thomas J. Ayres, Vladimir A. Babeshko, Bradley Paul Barber, Terence J. Bazow, Jonathan W. Benson, Jean-Paul Andre Berhault, Kaj G. Bodlund, Gary E. J. Bold, Newell O. Booth, Martin Bouchard, Nicholas J. Boulter, Nadejda I. Bouraou, Raymond K. Chow, David E. Coate, Marion F. Cohen, Therese Deierlein, Robert W. Derrow, Leonid M. Gelman, Robert H. Gilkey, Allen L. Gorin, John J. Gray, Ferdinand W. Grosveld, Michael Grunewald, Nail A. Gumerov, Robert V. Harrison, Mary C. Hofbeck, Richard H. Hoops, Stewart H. Hulse, Takehiro Isei, Don H. Johnson, Peter W. Jusczyk, Kelly Q. Kay, Ruth E. Keenan, Richmond G. Kent, Ronald T. Kessel, Kwang Yul Kim, Walter F. King, Shigeru Kiritani, Dana L. Kirkegaard, J. M. Klosner, Jun Kusakari, Lori F. Lamel, Robert A. Larabell, Zouhair Lazreq, Stephen G. Lindsey,

Guanghua Liu, Jeffrey J. Loether, I-Tai Lu, Richard W. Lungstrum, German A. Maksimov, Sebastien Manneville, Timothy J. Mason, Sven L. Mattys, E. Glenn McLeroy, Felipe Montoya-Moreno, William G. Morrison, Aly H. A. Moustafa, Joseph W. Myatt, Armand A. Nejade, Charles T. K. Ng, Arnold W. Novick, James C. P. O'Connor, Frederic M. Oran, Maria Pavlidou, Daniel A. Quinlan, Harry D. Record, Robert W. Rieder, Yuriy A. Rossikhin, Hiroyuki Sakai, Ernst Schonthal, Alan R. Selfridge, Eugenii L. Shenderov, Steven R. Sorenson, Thomas N. Stein, David Strelioff, Andrew C. Tam, Vivien C. Tartter, Aaron R. Thornton, Brian J. Timerson, Sushil Kumar Tomar, Ronald L. Underwood, Rajendran Venkatachalam, Andrew J. Vermiglio, George M. Vermillion, Dennis C. Walton, James J. Welsh, Peta J. White, Colman Wong, Pierre Zakarauskas, Chunnan Zhou, Robert F. de La Croix—*Members*

Lars N. Andersen, Leonor Aronson, Rodolfo T. Arrieta, Jonathan F. Ashmore, Harry J. Auvermann, Joan C. Bachenko, Frank A. Baffoni, Timothy J. Barnhart, John R. Baylis, Sean M. Beckman, Thomas K. Berger, Arthur Bettega-Correa, Carlos A. Bidondo, Jeremy J. Bielecki, Diane S. Bishop, Lawrence E. Blakely, Donald B. Bliss, Kevin P. Bongiovanni, Ben Bonham, Steven G. Borchardt, Thomas E. Bordley, Clyde C. Boylls, Arthur M. B. Braga, Karl W. Brandmaier, Larry R. Brown, Stephan V. Bugaj, Anthony T. Cacace, David A. Cartes, Jeff M. Cerjan, Marjorie K. M. Chan, Pascal Chedeville, Roger D. Chelf, John C. Clemens, K. A. Coleman, Robert J. Collum, Jeremy G. Colman, Dean G. Constantine, Leigh Cowan, Anna M. Crawford, Ben H. Crockett, Laurence Crosby, Matthew P. Daily, Michael Dallal, Arthur L. Daniel, Jeffery S. Davis, Lloyd A. Dawe, Emile De Kleine, Sergio De Rosa, Jeffrey J. DeMoss, Craig A. Devin, Diane E. DiMassa, Solon Diniz, John J. Ditri, Patricia A. Dorn, Susan M. Dumbacher, Alexandre M. Dykhne, David B. Elledge, Marta I. Etcheverry de Milou, Justin W. A. Fackrell, Wen-Tsun Fang, Philip P. Faraci, Glen E. Farrell, Maria Angela G. Feitosa, Michael A. Fisher, Alison B. Flatau, Shawn Gao, William R. Garmer, Laurie F. Garrison, David L. Gillette, Jose C. Giner, Jeffery A. Giordano, Paul M. Goggans, Mary C. Gruszka, Sung K. Ha, Benjamin Halberstam, Neil L. Halpern, Miriam Hamilton, Amin M. Hanafy, Andreas E. Hedqvist, Pedro Hernandez-Canedo, James P. Holm, John F. Holzrichter, Kiyoshi Honda, Kin L. Hong, Thomas B. Huffman, Allen J. Hundley, Lisa L. Hunter, John J. Iredale, Shrinivas N. Iyer, Jaehee Jang, Peter M. Janker, Raymond A. Janssen, Jennifer A. Johnson, Christopher D. Jones, Ajay Kapur, Janis Karklins, Prasanna Karpur, Tae-Gun Kim, Charles B. King, Karen IlerKirk, Leon Kloostra, Leah M. Knightly, Ben H. M. Kok, Robert P. Kraus, Sridhar Krishnaswamy, Theodore J. Kuemmel, Derek C. Lang, William Lang, Robert C. Lange, Hugh G. Larsen, John L. Lawless, Ricardo O. Lazzari, Paul A. Lepper, Martin Lesser, Thomas N. Lewis, Gau-Feng Lin, Donald Liu, Kenneth R. Lohr, Ya Yan Lu, Robert S. Luttrell, Michael W. Macon, Brian F. Magill, Galina V. Maralina, Robert H. Margolis, Olga M. Martynenko Kresse, Keener D. McClelland, Jacques Mehler, Nicole S. Metzger, Adele W. Miccio, Trevor A. Middel, Michael L. Mimnagh, Charles L. Monjo, T. C. Montgomery, Sherrill R. Morris, Ivano Morrone, Ami Moyal, Dan L. Nelson, Kin Wah Ng, Dorie A. Nojola, Thomas Norris, Ronald F. Norwood, John O'Connell, Roger M. Oba, Peter Olsson, Jane M. Opie, M. Serhat Ozyar, Ted Painter, Christopher A. Papadimos, Steve P. Parry, Heliton L. Paulini, Peter O. Paulson, Richard A. Pawlowicz, Felix A. Pena, Rafael P. Perez, Terry D. Plemons, Rodolfo Politano, Peter S. Popolo, Gopu R. Potty, Daniel Pressnitzer, Joseph C. Rampone, Jennifer G. Rasimas, Satha T. Raveendra, Charles Read, Benjamin K. Reaves, Christopher M. Richards, Maritza Riera-Gaxiola, Harold C. Robinson, Gonzalo J. Rodriguez, Marcelo Rosa, David Ross, Dan R. Salden, John R. Salisbury, William E. Sams, Nesrin Sarigul-Klijn, Christopher A. Saunders, Ramon F. Schapley, Hans P. Schedl, Natalie Schilling-Estes, Stephen C. Schreppler, Aleksander P. Sek, Holli B. Shan, Ming J. Sheu, William Shi, Syed B. Shuttari, Carol A. Simpson, Dan Simpson, Randy W. Smith, Scott W. Smith, Maria J. Sole, M. Spivack, Janet W. Stack, Gary Starr, Karl E. Sundkvist, Chia-Chi Sung, Richard P. Szwerc, Patrick Tamburro, Lucio N. Tello, Eric J. Terrill, Denis Thenail, Brian C. Thompson, Juan A. Torres, Kelly L. Tremblay, Giovanni Tristano, Renetta G. Tull, Martine Turgeon, Stefan J. Turneaure, Nancy Tye-Murray, Christine Valle, Nancy J. VanDerveer, Jean H. M. VRm.en, Gregory H. Wakefield, Amanda C. Walley, Kiiinon Wangpattanamongkol, Darren B. Ward, Nigel J. Wattrus, John Wendler, Emily R. Wilkinson, Barry S. Williams, Robert C. Williamson, Robert S. Wilson-DesRochers, James H. Wyatt, Quing Zhu, Robert A. Zurek, Elizabeth K. von Muggenthaler—*Associates*

Carla L. Adams, Visarut Ahkuputra, Sabri Ahmad, Jeffrey S. Allen, Troy M. Andersen, Michael D. Anderson, Kristen A. Armstrong, Rohima Badri, Paul T. Barthmaier, Adam Q. Bauer, Ravinder S. Beniwal, Zank Bennett, Peter M. Berkhout, James E. Bishop, David B. Blakeman, Arthur Blanc, Thomas C. Bonde, Alyssa C. Bonnoit, Jeff P. Boyczuk, Jodi P. Bray, Marc A. Brennan, Nashlea A. Brogan, Eric A. Broughton, Shannon T. Browne, Frank Louis Bruton, Zeeky A. Bukhala, Trevor A. Butler, John J. Buysman, Gregory S. Campbell, Jose M. Carmena, Lucy M. Carpendale, Henry Carron, Barton P. Carter, Ali T. Cemgil, Gabriel E. Chaparro, Brett J. Chigoy, Paul A. Dayton, Steven K. de la Vaux, Sharin R. de Silva, Adam T. Devlin, Roberto M. Dizon, Witold Dubaniewicz, Joshua J. Duncan, Jason M. Dvelis, Marie E. Ebbing, Sarah J. Eddy, Barrie A. Edmonds, Michael J. Epstein, Roman G. Eskue, Michele E. Evans, Yun-Hui Fan, Reif Faulkner, Dong Fei, Kristen M. Fishbeck, Tracy S. Fitzgerald, Cecile Fougeron, Michael J. Freeburg, Sara E. Gehr, Christopher M. Gibbons, Rene Headrick Gifford, Bernd Gottinger, Cristel Granat, George F. Griffo, Roger S. Grinnip, Oneil W. Guthrie, Jason R. Haaheim, Phillip M. Hamilton, Muhammad Haroon, Ryan B. Harvey, Anthony M. Hasenour, Wes S. Haveman, Genevieve M. Haviland, Zachary M. Helms, Marcia K. Hoelting, Gustavo A. Hoffmann, Mark S. Honea, Erin M. Horan, Christian A. Hresko, Jun Huang, Muhammad Z. Ikram, Nandini Iyer, David W. James, Wayne Jerzak, Athena C. M. Johns, Elizabeth K. Johnson, Krista L. Johnson, Iain T. Johnstone, Katerina Kaouri, Joe G. Keranen, Byung-chul Kim, Cecilia J. Kirk, Jonathan P. Klaas, Ryan Knox, Alexei Kochetov, Jason M. Koski, Michael A. Kraus, Tara R. Lance, Jennifer L. Lantz, Daniel F. Largen, Byung Kwon Lee, Sunghoon Lee, Young A. Lee, Brian R. Lemon, Taryn M. Leonard, Tang-Chun Li, Leonid M. Litvak, Li Liu, Daniel Bruce Lockhart, Manuel M. Mallebrera, Matthew D. McMahon, Kathleen M. Measer, David S. Michaud, Tricia A. Missall, Jeffery R. Moffitt, Sophie Molholm, Xiaomin Mou, Laura C. Muller, Jennifer A. Munro, Brian M. Napoletano, Coffey Neil, Patrick J. Nelson, Allen B. Nock, Benjamin E. Noel, Ryan P. O'Halloran, Christopher O. O'Malley, Miguel A. Obando, Sunyoung Oh, Tetsuro Oishi, John E. Orgar, Shadi F. Othman, Bradley D. Palmer, Hee-Suk Pang, Tomas S. Paredes, Christopher D. Park, Duk W. Park, Junhong Park, Sewon Park, Paavo Parn, Pedersen J. Pedersen, Martin E. Pelletier, Yvan Pelletier, Hernan E. Perez, Susan R. Perry, Kenneth J. Pfau, Madelaine C. Plauche, Sandra L. Poliachik, Simon T. Pranaitis, Scott S. Quick, Elizabeth A. Rickter, Hank S. Rinehart, Joshua L. Roby, Kendra L. Russell, Kazi Saifuddin, Terri L. Schneider, Kali A. Seney, Patricia J. Seymour, Yijun Shi, Kazushi Shimizu, Jinho Shin, Carlos A. Silva, Angela L. Slama, Kevin Smet, Alan K. Smith, William Soares-Filho, Amanda L. Spencer, Scott Spiegelberg, Catriona M. Steele, Julie K. Stucky, Lei Sun, Shinichi Takehara, Lawrence S. Taylor, Christopher J. Tebbens, Charles D. Theurer, Liau Y. C. Thomas, Vikrant Tiwari, Angus C. Tong, Giovanni Torelli, Paskorn Treesuwan, Shelli J. Trower, Fatma Nazan Ucar, Andrew C. Verneuil, Peter D. Viechnicki, Willis J. Warren, Chad D. Weltzin, Carl S. Werner, Ross A. White, Julie L. Whitten, Filip W. Windels, Patrick C. M. Wong, Lei Wu, Adam J. Wunderlich, Geun-Tae Yim, Steven J. Younghouse, Alan C. L. Yu, GongQiang Yu, Jong-Hwa Yu, Ahmad Zakaria, Richard S. Zelinger, Chao Zhang, Hua Zheng, Wei Zou—*Students*

Fellows	890
Members	2496
Associates	2646
Electronic Associates	170
Students	866
Total	7068

REPORTS OF RELATED MEETINGS

This Journal department provides concise reports of meetings that have been held by other organizations concerned with acoustical subjects; and of meetings co-sponsored by the Acoustical Society but planned primarily by other co-sponsors.

2002 ASME International Mechanical Engineering Congress and Exposition

The 2002 ASME International Mechanical Engineering Congress and Exposition (IMECE) took place in New Orleans, Louisiana in November 2002. During this congress, the Noise Control and Acoustics (NCA) Division sponsored a total of 14 sessions with 11 symposium sessions and 3 special sessions that included (1) the *Fifth International Symposium on Fluid-Structure Interaction, Aeroelasticity, and Flow-Induced Vibration and Noise*; (2) *Recent Advances in Active Noise Control*; (3) *Active Noise Control in Transportation Systems*; (4) *Methods for Inverse Acoustic Problems*; and (5) *Vibrations in Piping Systems*. These symposia contained 45 papers by 90 different authors. The highlight of these sessions was the Rayleigh Lecture, given by Dr. Ilene Busch-Vishniac, Dean of the Whiting School of Engineering, Johns Hopkins University, on ''The Big Problems Remaining in Transduction.'' Her presentation was quite unconventional and the following discussions were very thought provoking. The 2002 tutorial lecture was presented by Dr. John Fahnline from Penn State University, who gave an excellent background on boundary element techniques and radiated power computation. As last year, the NCA Division organized a special panel session to discuss issues in noise control. Each of the technical committee chairs made a presentation on the emergent issues and developments in their specific technical areas. This panel discussion session also provided a forum for the mechanical engineering community at large to share their concern and interest with the technical committees prior to their planning sessions for the future IMECEs. Finally, the 2002 NCAD Best Paper Award selected by the Executive Committee went to Jason M. Anderson, Ricardo Burdisso, and Wing Ng of the Vibration and Acoustic Laboratories, Department of Mechanical Engineering, Virginia Polytechnic University, for their paper entitled, ''An Active Flow Distortion Control System for Serpentine Inlets.''

SEAN F. WU

 0001-4966/2003/114(1)/17/1/$19.00

BOOK REVIEWS

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

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

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

Higher-Order Numerical Methods for Transient Wave Equations

Gary C. Cohen

Springer-Verlag, New York, 2001.
348 pp. Price: $69.95 (hardcover) ISBN: 354041598X.

The book *Higher-Order Numerical Methods for Transient Wave Equations* written by Dr. Gary C. Cohen is concerned with the numerical methods for transient acoustic, elastic, and electromagnetic waves, although the acoustic wave equations receive the most comprehensive treatment. The book introduces higher-order finite difference and finite element methods for the time-dependent wave equations. Such time-domain methods have important advantages over the frequency-domain methods in the wideband simulation and nonlinear problems, although the book is concerned with linear wave equations. The author is exceptionally well qualified to write such a book because he and his collaborators at INRIA and elsewhere have contributed to many fundamental materials covered by this book.

The book is comprised of three parts: The first part deals with the basic wave equations, definitions, and properties; the second part introduces higher-order finite difference methods; and the third part concentrates on the mass-lumping finite element methods.

In the first part, Chap. 1 introduces the basic acoustic, elastic, and electromagnetic wave equations and boundary conditions. Chapter 2 summarizes the essential basics for functional spaces, variational formulations, and energy identities. In Chap. 3 the plane wave solutions are given for the homogeneous wave equations.

In Part II, the first four chapters (Chaps. 4–7) are concerned with the finite difference methods for homogeneous media, where the important issues including the dispersion relation, stability, numerical dispersion, and anisotropy of the methods are discussed. Second-order and higher-order approximations are used for both spatial and temporal discretization. Chapters 8 and 9 are concerned with the finite difference methods for heterogeneous media, where the stability of the schemes is analyzed by the energy techniques. In Chap. 10, the reflection and transmission of waves at a planar interface between two materials are investigated to analyze the accuracy of the finite difference methods in discontinuous materials. These analyses highlight the difficulties of the finite difference methods for discontinuous media, especially for those with complex geometries.

As a natural alternative to the finite difference methods for complex geometries, the finite element methods are developed in Part III. The novel idea of mass lumping is first developed in one dimension in Chap. 11, to eliminate the need to invert a mass matrix in the conventional finite element methods. This idea is extended in Chap. 12 to 2D and 3D scalar wave equations with quadrilateral and hexahedral finite elements, respectively, where the straightforward tensor products are used to arrive at the 2D and 3D grids within each element. Lower-order mass-lumping schemes are also developed for triangular and tetrahedral elements. However, the book does not present such schemes for orders higher than 3, and it is not clear how the method can be extended to such high orders. Furthermore, for 3D problems, the cost of mass lumping method seems very high, even for lower-order schemes. Chapter 13 extends the mass-lumping idea to mixed formulations and edge elements that are more appropriate for vector (such as electromagnetic and elastic) wave equations. Again, the quadrilateral and hexahedral elements prove to be much more efficient than triangular and tetrahedral elements. Chapter 14 is devoted to a new absorbing boundary condition, the perfectly matched layers (PMLs), for the truncation of unbounded media.

This book represents some interesting research ideas from the author's group. In connection with both higher-order finite difference and finite element methods discussed in this book, I think it is worthwhile to mention a few related recent advances. In the arena of finite difference methods, it is well known, also as pointed out in the book, that complex geometries are problematic because the inherent staircasing approximation will greatly reduce the accuracy. However, several recent works proposed embedded interfaces in the finite-difference methods to match the boundary conditions on these interfaces to ensure a global second-order and fourth-order accuracy for the second- and fourth-order finite difference methods, respectively. These methods potentially can greatly improve the accuracy of finite difference methods for complex geometries.

In the area of higher-order finite element methods, some related new methods are the discontinuous Galerkin's methods, spectral and pseudospectral time-domain methods. These methods do not use edge elements, but discontinuous elements of higher orders, together with the Riemann solver or penalty method for making the flux correction so that boundary conditions are enforced weakly between adjacent elements. These new methods seem to be very promising alternatives of the higher-order finite element methods presented in this book.

In summary, this book is a very valuable reference for the readers of this journal who are interested in the computational methods for transient waves. What make this book unique are the novel mass-lumping techniques for finite element methods, which are still being actively investigated.

QING HUO LIU
Duke University
Department of Electrical and Computer Engineering
Box 90291
Durham, North Carolina 27708

The Science and Applications of Acoustics

Daniel R. Raichel

Springer-Verlag, New York, 2000.
598 pp. Price: $69.95 (hardcover) ISBN: 0387989072.

The tenth volume in the distinguished AIP series Modern Acoustics and Signal Processing is somewhat of an exception in that it does not deal with a very limited topic within the broad field, but rather presents us with a text, usable in the classroom, surveying that broad field. While coming to us as a survey, it nevertheless prepares us with a thorough and mathematically rigorous theoretical basis before embarking on the journey through the acoustical archipelago.

Chapter 1 presents a marvelously concise, yet quite exhaustive, review of the history of acoustics from its roots including ancient Chinese bells to current experiments in sonoluminescence.

After a brief conceptual opening Chap. 2 introduces us in rather sophisticated mathematical rigor to important acoustics equations. The problems given at the end of the chapter not only apply basic concepts of the chapter, but insist on some fundamental understanding in order to apply the given information to the formulas derived in the text.

 0001-4966/2003/114(1)/21/2/$19.00

Chapter 3 presents concepts such as basic wave propagation, complex waves, standing waves, Huygen's principle, the Doppler effect, reflection, refraction, diffraction, octave and one-third octave bands, sound pressure level, weighting curves, performance indices for environmental noise, particle displacement and velocity, correlated and uncorrelated sound, intensity, the monopole source, spherical and hemispherical waves, and energy density. The introduction of almost all of these includes practical examples with relevant problems and solutions.

The discussion of vibrating strings in Chap. 4 includes the development of the transverse wave equation and solutions with the imposition of boundary and initial conditions, as well as consideration of forced vibrations.

Vibrating bars are treated in Chap. 5 while considering both longitudinal and transverse waves, with varying boundary conditions.

The natural extension to two dimensions is found in Chap. 6, which considers the stretched membrane with examples of a number of different boundary conditions, including an application to the membrane of a kettle drum. This chapter also outlines an approach to the thin plate problem.

Chapter 7 treats sound waves in enclosed spaces with emphasis on pipes, waveguides, and resonators.

The subsequent Chap. 8 concludes the more introductory general chapters by treating acoustic analogs, ducts, and filters.

In the remaining ten chapters the text becomes more topically oriented by treating sound measurement instrumentation, hearing, and psychoacoustics, with two chapters devoted to architectural acoustics, two chapters to noise control, and one chapter each to underwater acoustics, ultrasonics, music and musical instruments, and vibration and vibration control.

In such an all-encompassing text, the depth of several individual topics covered at an introductory graduate level is truly astounding. At the same time, in most cases, details which are not addressed in the text can be found in the relatively complete set of references given at the end of each chapter. There are some notable omissions, which, due to the survey character of the text are understandable, yet inclusion or at least a suggestion of awareness would enhance the value of the text. As examples, in the music chapter no mention is made of problems inherent in tuning, nor of the characteristic piano sound resulting from inharmonic contributions to the stiff piano string spectrum, or of the challenge presented to makers of percussion instruments to design a structure which favors enhancement of harmonic overtones, when the transient excitation does not require overtone harmonicity.

In summary, this is a remarkable text at the advanced undergraduate or introductory graduate level. Its strength lies in its breadth of coverage, in its level of sophistication, and in its didactic value. Many concepts introduced in the chapters are illustrated by applied example problems with solutions. Additional problems at the end of each chapter are carefully chosen to provide practical computational exercises while demanding conceptual understanding. References include relevant texts, summary monographs, and current journal articles. The structure of the second half of the book as relatively independent chapters lends itself nicely to course construction. Chapters with less relevance to a particular field can easily be omitted and others can readily be enhanced with the use of the reference section at the end of each chapter.

UWE J. Hansen
Department of Physics
Indiana State University
Terre Haute, Indiana 47809

 0001-4966/2003/114(1)/22/2/$19.00

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, Univ. of Rochester, Rochester, New York 14627*

6,510,107

43.30.Sf ACOUSTIC METHOD AND SYSTEM FOR MEASURING FISH POPULATION DATA IN LITTORAL ENVIRONMENTS

Orest Diachok, assignor to the United States of America as represented by the Secretary of the Navy
21 January 2003 (Class 367/131); filed 9 March 2001

A system for determining fish population data such as numbers, sizes, and temporal characteristics consists of readily deployable and retrievable source and receiver units. These units are bottom anchored and suspended from a buoyancy member. The anchors can be jettisoned so that the units will rise to the surface for recovery. It is preferable to deploy a source unit and a receiver unit both above and below the thermocline, assuming such exists. A source unit contains power supply and controller. A receiver unit contains up to four hydrophones, power supply, signal conditioning means, A-D converter, computer, and memory storage unit.—WT

6,501,704

43.30.Xm UNDERWATER OBJECT POSITIONING SYSTEM

Kiyokazu Nishimura, assignor to National Institute of Advanced Industrial Science and Technology; Kiyokazu Nishimura
31 December 2002 (Class 367/130); filed in Japan 22 November 2000

A system for locating and positioning an actively communicating underwater object comprises a surface workboat with an acoustic transmitter-receiver unit and two towed surface units, the first of which supports an acoustic transducer while the second supports two acoustic receivers. Both tow units are connected to the workboat by a communication cable and their spatial positions are determined by the GPS. An acoustic pulse transmitter is mounted on the underwater object. Its position in then determined from various sound travel time measurements.—WT

6,512,980

43.30.Xm NOISE REFERENCE SENSOR FOR USE IN A DUAL SENSOR TOWED STREAMER

Frederick J. Barr, Pearland, Texas
28 January 2003 (Class 702/1); filed 19 October 1999

A towable seismic streamer features three sensors. The first generates a signal due to the acoustic pressure in the surrounding water, the second a signal due to the vertical component of particle velocity in the water, and the third a signal indicative of the vertical motion of the streamer itself relative to the water. By combining these three signals in various ways, a substantially noise free signal is obtained, which is not corrupted by surface reflections.—WT

6,449,593

43.38.Hz METHOD AND SYSTEM FOR TRACKING HUMAN SPEAKERS

Päivi Valve, assignor to Nokia Mobile Phones Limited
10 September 2002 (Class 704/233); filed 13 January 2000

This microphone array is intended for locating and tracking multiple speakers in an automobile. Processing of the mic signals is fairly typical of a beamformer, with options for either filter-and-sum or delay-and-sum. The basic beam steering criterion appears to be based on a comparison of the overall power levels of multiple beam signals after beamformer processing. Beam detection is not continuous but repeats in a cycle of approximately 0.1 s. Overall beam gain can then be individually controlled for each speaker.—DLR

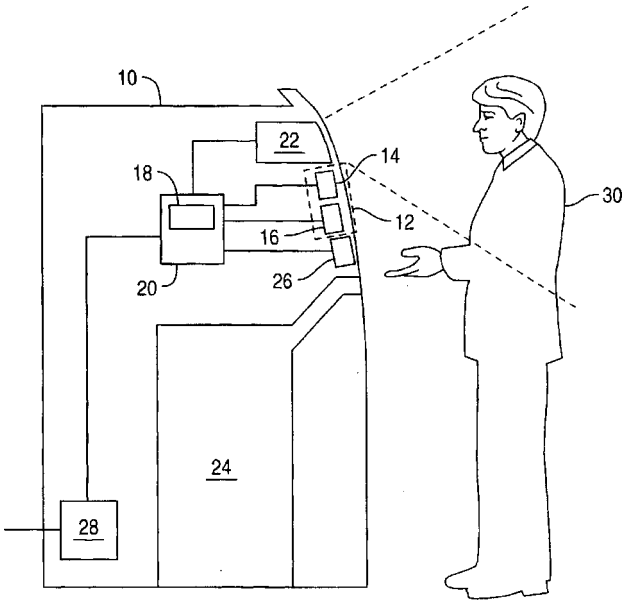
6,494,363

43.38.Hz SELF-SERVICE TERMINAL

Ian M. Roger *et al.*, assignors to NCR Corporation
17 December 2002 (Class 235/379); filed 13 January 2000

Contemporary teleconferencing systems may incorporate steerable or selectable microphones, steerable or selectable loudspeakers, and a means to accurately locate individual participants. What seems to be patented here is

 0001-4966/2003/114(1)/23/19/$19.00



the idea of using this technology in a self-service terminal such as an automatic bank teller. In an ideal world the machine would be able to focus sound pickup and transmission on a specific zone of coverage surrounding the targeted user's head. The patent claims are quite emphatic about this: ''...wherein only the user within the privacy zone is capable of hearing messages directed to the user... .'' In contrast, what can actually be achieved in practice may be somewhat disappointing.—GLA

6,506,154

43.38.Hz SYSTEMS AND METHODS FOR CONTROLLING A PHASED ARRAY FOCUSED ULTRASOUND SYSTEM

Avner Ezion *et al.*, assignors to InSightec-TxSonics, Limited
14 January 2003 (Class 600/437); filed 28 November 2000

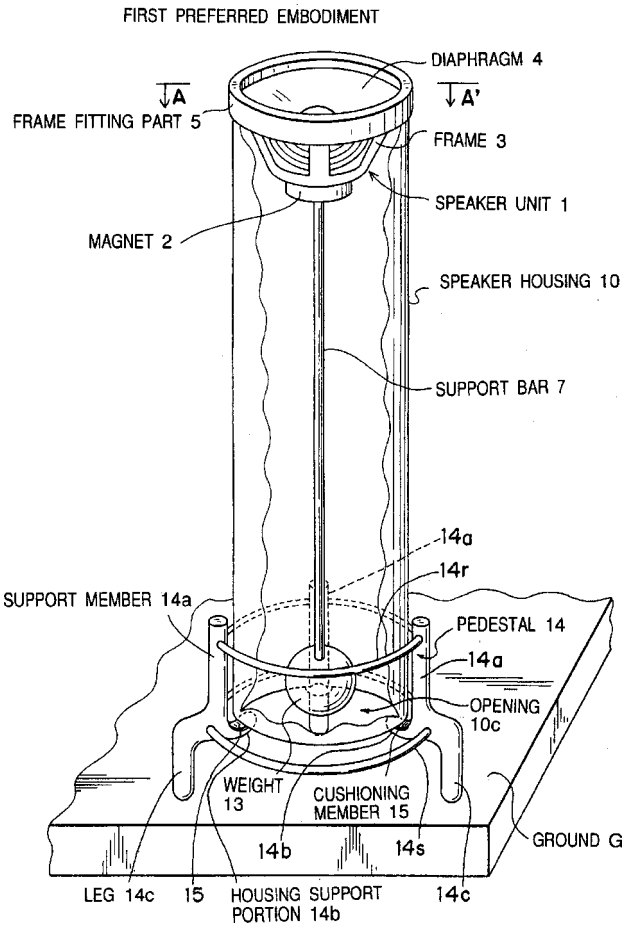
The patent describes a system and method for controlling a phased array transducer in a focused ultrasound system. Digital controls are used to scale the phase and amplitude of three or more base sinusoids which are linearly combined into component sine vectors. Digital controls are also provided for switching between various distances, shapes, and orientations of the focal zone of the transducer.—DRR

6,484,843

43.38.Ja SPEAKER APPARATUS EQUIPPED WITH MEANS FOR PRODUCING COMPLICATED WAVEFORM OF LOW FREQUENCY WITH HIGHER IMPROVED FIDELITY

Hiroyuki Yoshii and Haruhiko Imamura, assignors to Timedomain Corporation
26 November 2002 (Class 181/153); filed in Japan 2 December 1999

A heavy pendulum **7**, **13** is attached to magnet **2** of upward-firing



loudspeaker **1**. Its purpose is to suppress unwanted vibrations of the magnetic assembly.—GLA

6,485,241

43.38.Ja SURFACE MOUNT RING ASSEMBLY FOR LOUDSPEAKER

J. Craig Oxford, Nashville, Tennessee
26 November 2002 (Class 411/339); filed 10 June 1997

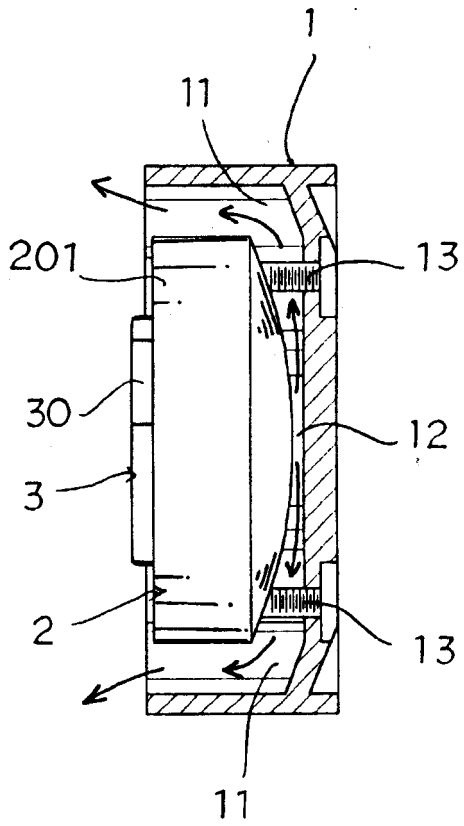
Anyone who owns a pickup truck knows that the rear window is completely supported by its mounting gasket. It is simply snapped in place, requiring no screws, brackets, or other fasteners. Well, hey, why not use the same trick for wall-mounted loudspeakers?—GLA

6,487,297

43.38.Ja STRUCTURE OF A LOUDSPEAKER

Yen Chin-Fa, Taipei, Taiwan, Province of China
26 November 2002 (Class 381/345); filed 22 November 2000

This high-frequency compression driver requires no horn. Instead, sound energy generated by diaphragm **12** travels outwardly through ''reso-



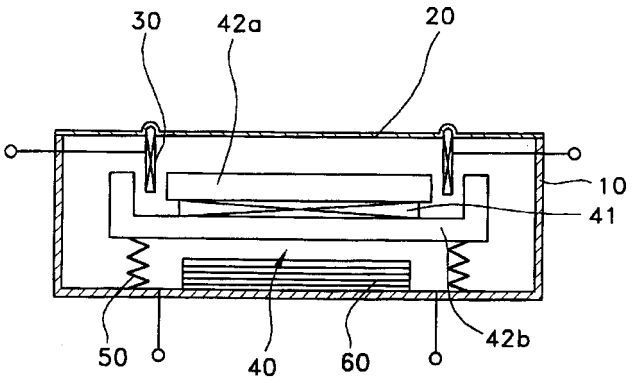
nance chambers'' **11** and emerges from what is effectively a ring-shaped slot. The resulting sound quality is said to be ''steady and loud.''—GLA

6,487,300

43.38.Ja VIBRATION SPEAKER

In Ho Lee *et al.*, assignors to Samsung Electro-Mechanics Company, Limited
26 November 2002 (Class 381/396); filed in the Republic of Korea 17 December 1999

This pager transducer has separate voice coils for two modes of operation. Coil **30** drives case **20** to produce sound. Coil **60** vibrates the entire



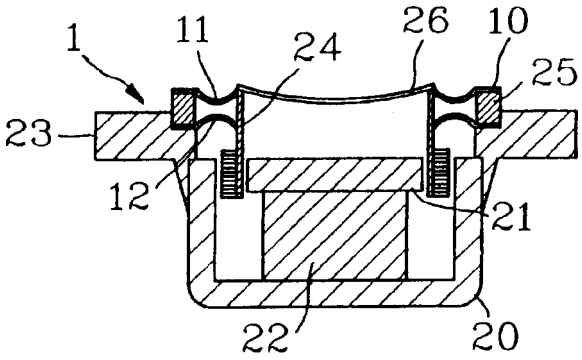
spring-mounted magnetic assembly **40** to produce a strong, tactile signal.—GLA

6,490,363

43.38.Ja STRUCTURE OF SPEAKER

Chun-I Liu, Taoyuan City, Taiwan, Province of China
3 December 2002 (Class 381/403); filed 13 October 1999

The edge-driven moving assembly **26** of this loudspeaker is supported by compliant surrounds **11** and **12**, which are spaced far enough apart to



resist rocking modes. The patent claims are quite specific in describing the geometry shown—the diaphragm itself is concave and the two suspension members have oppositely facing curved surfaces.—GLA

6,493,452

43.38.Ja SPEAKER UNIT

Hiroshi Koizumi *et al.*, assignors to Sony Corporation
10 December 2002 (Class 381/182); filed in Japan 20 May 1998

Is it really possible that until now no one has mounted an oval tweeter in front of an oval woofer to form an oval coaxial speaker assembly?—GLA

6,493,456

43.38.Ja THIN SPEAKER ASSEMBLIES INCLUDING LATERALLY OFFSET RESONATOR CAVITIES AND PERSONAL ELECTRONIC DEVICES INCLUDING THE SAME

Magnus F. Hansson, assignor to Telefonaktiebolaget L. M. Ericsson
10 December 2002 (Class 381/345); filed 18 October 2000

Cellular telephones become thinner and thinner. To achieve satisfactory sound quality from a very shallow speaker assembly this invention makes use of laterally disposed resonant cavities.—GLA

6,494,289

43.38.Ja DEVICE FOR DYNAMIC EXCITATION OF PANEL LOUDSPEAKERS

Wolfgang Bachmann *et al.*, assignors to Harman Audio Electronic Systems GmbH
17 December 2002 (Class 181/173); filed in Germany 15 May 1998

This patent describes an improvement on an earlier design. The driver of a panel-type loudspeaker is embedded within the thickness of the panel itself. To avoid making the panel more rigid and thereby inhibiting the travel of bending waves, the magnetic assembly is attached at only three points.—GLA

6,496,586

43.38.Ja THIN LOUDSPEAKER

Michael A. Hayes *et al.*, assignors to New Transducers Limited
17 December 2002 (Class 381/152); filed 10 December 1998

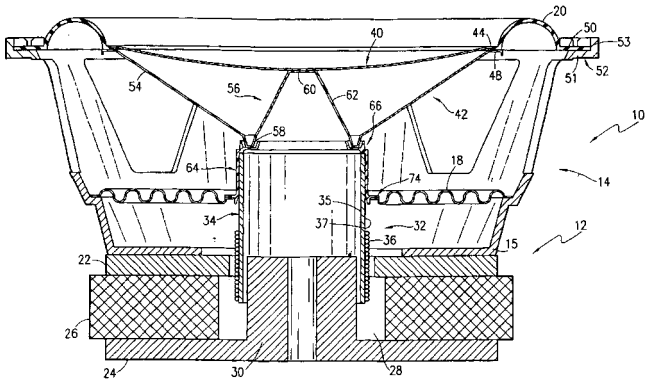
Mounting any kind of loudspeaker in a very shallow enclosure introduces all sorts of unforeseen problems. The patent describes a number of techniques intended to optimize the performance of a shallow loudspeaker system employing a combination of panel-form transducer elements and conventional cone-type loudspeakers.—GLA

6,496,590

43.38.Ja LOUDSPEAKER WITH IMPROVED DIAPHRAGM

Lucio Proni, assignor to JL Audio, Incorporated
17 December 2002 (Class 381/423); filed 8 December 2000

The double diaphragm geometry shown (**40**, **42**) is said to resist bend



ing and twisting at high excursions.—GLA

6,496,588

43.38.Kb DIRECTIONAL DYNAMIC MICROPHONE INTERCHANGEABLE TO HAVE UNIDIRECTIONAL AND SUPERDIRECTIONAL CHARACTERISTICS

Ching-Lu Chang, Taichung City, Taiwan, Province of China
17 December 2002 (Class 381/358); filed 11 March 1999

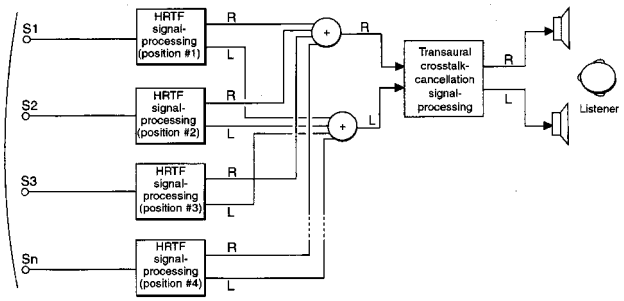
In 1944, Harry F. Olson demonstrated a ribbon microphone whose directional characteristics could be altered by a small sliding valve. This patent describes a dynamic microphone whose pattern can be changed between cardioid and hypercardioid by adjusting a rotary valve.—GLA

6,498,857

43.38.Vk METHOD OF SYNTHESIZING AN AUDIO SIGNAL

Alastair Sibbald, assignor to Central Research Laboratories Limited
24 December 2002 (Class 381/310); filed in the United Kingdom 20 June 1998

At least 50% of the world's audio engineers are currently working on ways to create a convincing three-dimensional sound field from two loud-



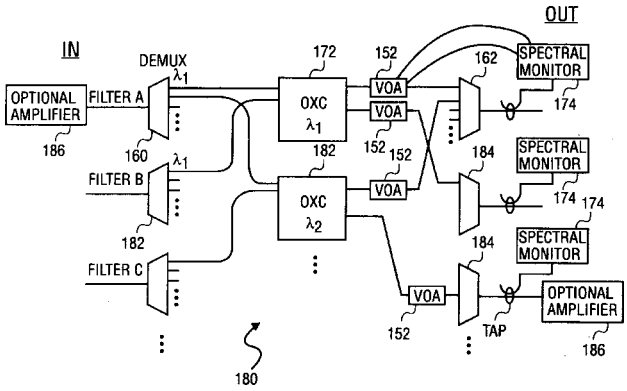
speakers. This patent discusses inherent problems in synthesizing near versus far sources and large versus small sources. It is clearly written and contains useful information.—GLA

6,510,261

43.38.Zp ACOUSTO-OPTIC VARIABLE ATTENUATOR WITH ACTIVE CANCELLATION OF BACK REFLECTIONS

Wayne V. Sorin *et al.*, assignors to Novera Optics, Incorporated
21 January 2003 (Class 385/27); filed in the Republic of Korea 6 June 1997

Based on previous United States patents 6,021,237 and 6,253,002, the current patent covers a variable optical attenuator (VOA), which is an all-fiber acousto-optical tunable intensity attenuator that includes an acoustic wave device to reduce backward-propagating sound waves. The acousto-optical filter includes an optical fiber divided into two regions. An acoustic wave generator coupled to the optical fiber produces an acoustic wave that



travels in a specific direction in one of the two regions. A backward-propagating wave is generated in response to the propagation of the acoustic wave along the optical fiber. An acoustic wave member is coupled to the optical fiber. Another acoustic wave generator coupled to the optical fiber in the second of the two regions produces another acoustic wave that combines with the backward-traveling acoustic wave, resulting in attenuation of the latter wave.—DRR

6,484,845

43.40.Vn METHOD AND SYSTEM FOR INFLUENCING POSSIBLE STRUCTURE-BORNE SOUND CONDUCTIONS AND POSSIBLE NOISE RADIATIONS OF OBJECTS

Ruediger Schleicher *et al.*, assignors to DaimlerChrysler AG
26 November 2002 (Class 181/207); filed in Germany 13 June 1998

The invention is an active gasket. Consider a situation, for example, in which structurally transmitted vibrations are coupled to a window via the glazing, which in turn acts as a diaphragm and produces airborne sound. It

is well known that this effect can be substantially attenuated by mounting the glass in a resilient gasket. If piezoelectric transducers could be imbedded in the gasket and driven by an antiphase signal, then further attenuation should be possible. The patent asserts that the idea is not only possible but practical.—GLA

6,509,081

43.50.Gf NO-SEPTUM ACOUSTIC SANDWICH PANEL, AND APPARATUS AND METHOD FOR SUPPRESSING NOISE IN A NOZZLE

John Anthony Diamond, assignor to The Boeing Company
21 January 2003 (Class 428/131); filed 28 September 2000

The panel comprises a honeycomb core with both sides covered by perforated metal and a sheet of metal cloth, such as woven wire or metal felt. The panels are arrayed in a radial pattern to suppress noise in a nozzle duct (such as a fan duct of a turbofan aircraft engine). The metal cloth sheets replace the need for a septum, and the patent claims that these panels are therefore significantly thinner and lighter than known panels.—CJR

6,503,187

43.50.Hg STIMULATOR HEAD AND METHOD FOR ATTENUATING THE SOUND EMITTED BY A STIMULATOR COIL

Risto Ilmoniemi, Espoo, Finland *et al.*
7 January 2003 (Class 600/14); filed in Finland 28 November 1997

A stimulator head is used to transduce electrical energy into magnetic energy that can be used to stimulate biological tissue, including the brain, the peripheral nervous system, and the heart through the application of an electromagnetic field composed of electric and magnetic fields. In order to generate the desired magnetic field, an alternating current must be applied to a coil, which can result in very rapid changes of force within the coil and thereby generate an acoustic wave in the range of 1–10 kHz with a SPL of 120 dB or above. The goal of the patented device is to attenuate that generated noise. This is achieved by housing the stimulator coil elements inside a structure which allows at least the stimulator coil itself to be enclosed by a shell. The space inside the shell is pumped to a vacuum.—DRR

6,507,790

43.50.Yw ACOUSTIC MONITOR

James V. Radomski, assignor to Horton, Incorporated
14 January 2003 (Class 702/39); filed 15 July 1998

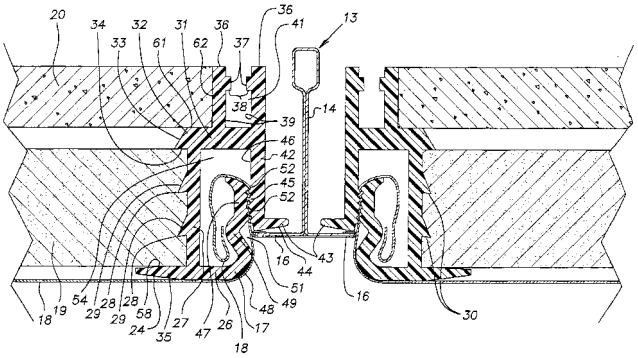
This is a self-contained system that detects faults in the operating condition of machinery by continuously analyzing the sound produced by the monitored source. The resulting power spectrum is compared to a previously recorded ''acoustic signature'' used as a baseline. In the preferred embodiment, the acoustic monitor performs a real-time $\frac{1}{12}$-oct digital bandpass filtering over an 8-oct range (midband frequencies from 33.108 to 8000 Hz) and computes the acoustic power output, in decibels, of each of the resulting 96 bandpass filters.—DRR

6,499,262

43.55.Ev CEILING PANEL

Mark A. Pinchot and Curtis P. Taylor, assignors to Frank Novak & Sons, Incorporated
31 December 2002 (Class 52/273); filed 11 September 2000

The suspended ceiling panel comprises a rigid frame, sound-absorbing material, sound attenuation material, and a face fabric. The fabric is secured



to the frame by gripping elements that are hidden by the frame. The edges of the frame and fabric are both hidden so the fabric seems to float evenly across the entire panel.—CJR

6,511,288

43.55.Vj TWO PIECE BLOWER HOUSING WITH VIBRATION ABSORBING BOTTOM PIECE AND MOUNTING FLANGES

William Stuart Gatley, Jr., assignor to Jakel Incorporated
28 January 2003 (Class 415/206); filed 31 August 2000

This blower draws in combustion gases for high-efficiency furnaces and propels the exhaust to the atmosphere. It has a housing with a resilient bottom piece and a rigid top piece. There is a cushion between the two to provide vibration isolation between the motor mounting plate and the furnace.—CJR

6,506,092

43.58.Wc METHOD AND APPARATUS FOR ENHANCING AN APPLAUSE

Thomas C. Kuracina and Randall E. Ohnemus, assignors to Intune Corporation
14 January 2003 (Class 446/81); filed 25 July 2000

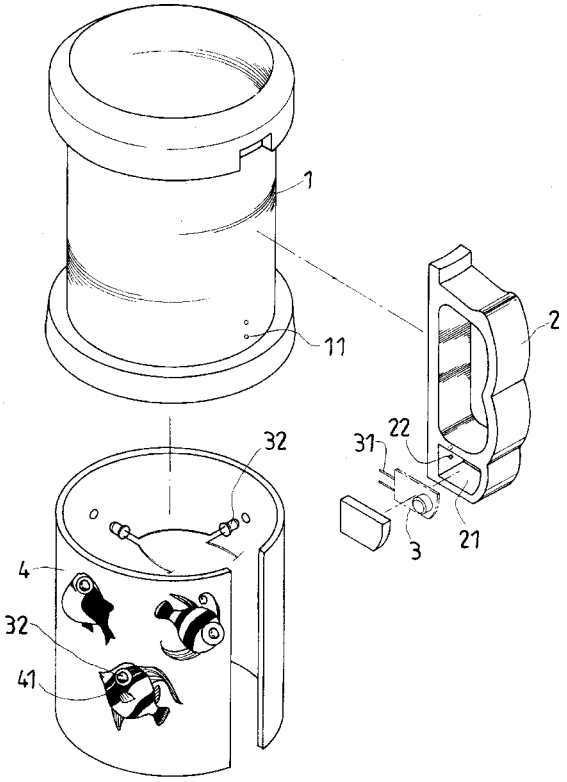
Fifty figures and ten claims later, it can be discovered that a beverage holder can be used to enhance applause by forming a horn.—MK

6,505,734

43.58.Wc CONTAINER WITH LIGHT AND SOUND EMITTING FUNCTION

Chin-Lai Su, assignor to Wea Bor Company, Limited
14 January 2003 (Class 206/217); filed 12 June 2001

Here is another entry in the audible beverage category. This time, it



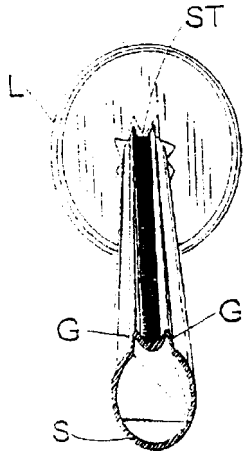
includes both light and a level sensor to detect tipping.—MK

6,512,727

43.58.Wc TALKING STRAW WITH GUIDE RIDGE

Marcus Garabedian, Orange, California
28 January 2003 (Class 369/63); filed 7 December 2001

Imagine a straw with a sound track **ST** that can be pulled across a



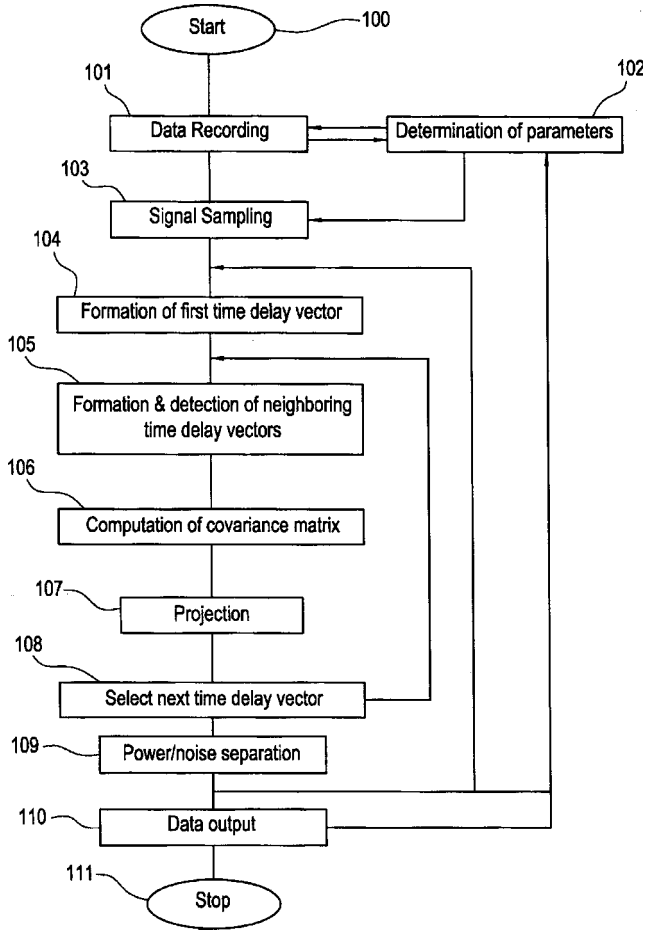
point (embedded in a lid) to play it back. Now, how long is this recording?—MK

6,502,067

43.60.Bf METHOD AND APPARATUS FOR PROCESSING NOISY SOUND SIGNALS

Rainer Hegger *et al.*, assignors to Max-Planck-Gesellschaft zur Forderung der Wissenschaften e.V.
31 December 2002 (Class 704/216); filed in Germany 21 December 1998

The *modus operandi* of the device is to record a nonstationary sound signal composed of power and noise components at such a rapid sampling rate that signal profiles within the observed sound signal contain sufficient redundancy for noise reduction. The redundancy is detected and correlations between signal profiles are determined within segments of the sound signal.



Correlated signal components are attributed to a power component and uncorrelated signal components to a noise component of the sound signal. These correlations between the signal profiles are established by the method of nonlinear noise reduction in deterministic systems in reconstructed vector spaces in the time domain.—DRR

6,507,819

43.60.Gk SOUND SIGNAL PROCESSOR FOR EXTRACTING SOUND SIGNALS FROM A COMPOSITE DIGITAL SOUND SIGNAL

Junji Yoshida *et al.*, assignors to Matsushita Electric Industrial Company, Limited
14 January 2003 (Class 704/500); filed in Japan 21 May 1996

This relates to a video and voice signal processing apparatus that outputs an intelligible voice signal even if frame decimation has been applied to the video signal. The apparatus receives an input containing encoded video and voice signals and recovers the encoded video and voice signals for output. A designated frame sequence is extracted from the encoded video signal, processed, and output as a representative video signal. The encoded voice signal is similarly processed and output as a voice signal.—DRR

6,446,001

43.66.Lj NAVIGATION APPARATUS FOR NAVIGATING A VEHICLE BASED ON SYMBOLIC SOUNDS

Hideki Sakai *et al.*, assignors to Sony Corporation
3 September 2002 (Class 701/211); filed in Japan 4 March 1999

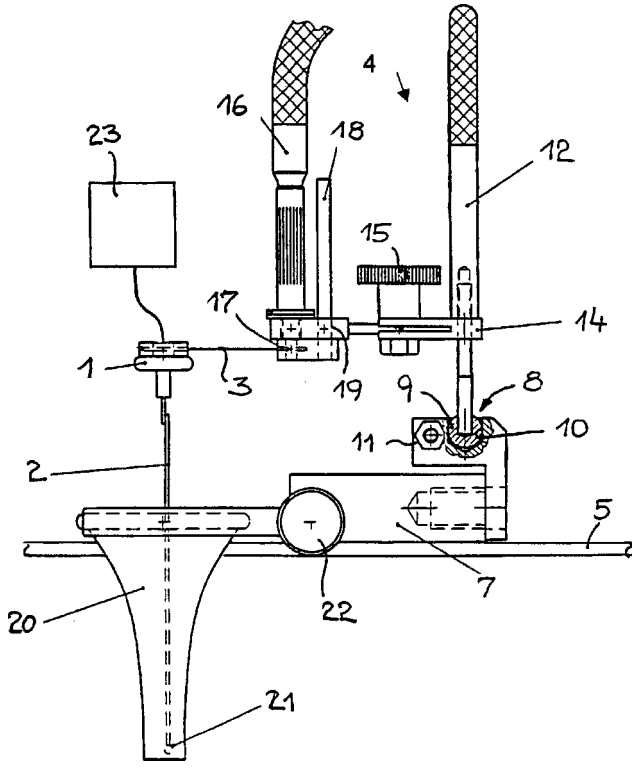
This automotive navigation device includes recorded environmental sounds appropriate for specific regions which may appear within the navigational range. Some cited examples include wave sounds to signal a nearby beach, bird sounds for a forest, or a gong if the vehicle is approaching a Japanese temple. The idea is to avoid distracting the driver's eyes from the road. The patent does not suggest sounds which might be suitable for a terrain feature or a field of crops.—DLR

6,506,164

43.66.Sr DEVICE FOR ELECTROMECHANICAL STIMULATION AND TESTING OF HEARING

Gabriele Reischl *et al.*, assignors to Phonak AG
14 January 2003 (Class 600/559); filed in Germany 22 March 2000

The patent covers the concept of a device for electromagnetic stimulation and testing of hearing. An electromechanical transducer produces mechanical vibrations which are mechanically coupled from the transducer



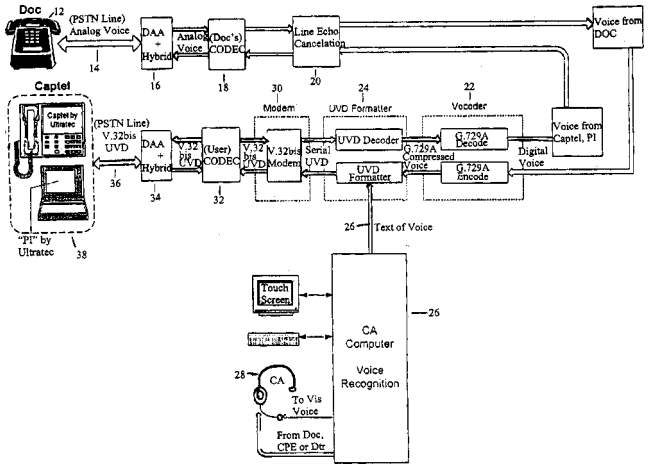
without surgery through the external auditory canal to the *umbo* and thus to the *manubrium mallei* of the ossicular chain. The coupling element is oriented with respect to the *umbo* and positioned by a secure, play-free linkage to the skull.—DRR

6,504,910

43.66.Ts VOICE AND TEXT TRANSMISSION SYSTEM

Robert Engelke, Madison, Wisconsin *et al.*
7 January 2003 (Class 379/52); filed 7 June 2001

The device constitutes a method for transmitting both sound and text over a telephone connection between a hearing user and an assisted user through a relay operator. The voice of the hearing user is digitized, creating a digital text at the relay corresponding to the spoken words. The digitized



voice and the text are combined into digital data packets, each of which includes a format control byte allowing the digitized voice data to be combined with the text of the words spoken by the hearing user. The combined packets are transmitted through telephone lines to the station of the assisted user so that the station can reconstitute both voice and text from the digital data packets.—DRR

6,510,206

43.66.Ts RELAY FOR PERSONAL INTERPRETER

Robert M. Engelke and Kevin Colwell, assignors to Ultratec, Incorporated
21 January 2003 (Class 379/52); filed 19 September 2001

This is yet another system to translate the spoken word into text to facilitate telephone communication between deaf people and hearing people. The system makes use of the TDD telephonic process commonly used by deaf persons. An operator, referred to as a ''call assistant,'' serves as a human intermediary between a hearing user and a deaf person. Normally the call assistant wears a headset that communicates by voice with the hearing user and also has access to a TDD device that can communicate with the

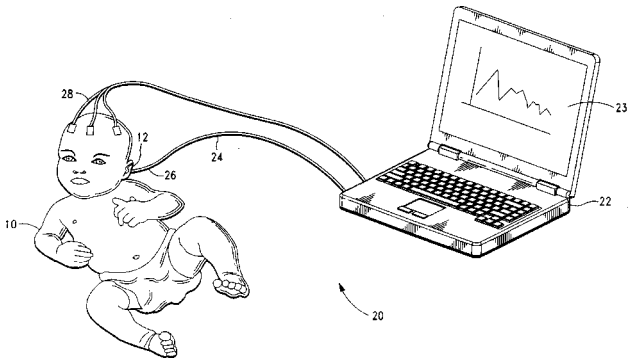
deaf user using a TDD protocol. To overcome the speed limitations inherent in typing, the call assistant does not type most words but instead revoices the words spoken by the hearing person into a computer operating a voice recognition software package trained to the voice of that call assistant. It is conceivable that an improvement in the voice recognition software may eventually lead to the demise of the call assistant, so that communication between the deaf person and the hearing person may be more direct and eliminate the need for intermediary typing.—DRR

6,503,207

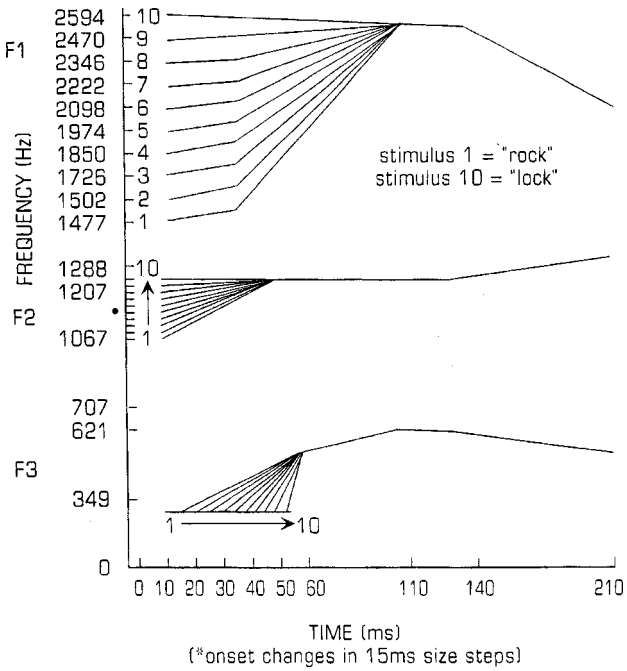
43.66.Yw MULTI-MODE AUDIOMETRIC DEVICE AND ASSOCIATED SCREENING METHOD

Robert T. Stone, assignor to Natus Medical, Incorporated
7 January 2003 (Class 600/559); filed 13 February 2001

This multimode audiometric device, aimed at testing the hearing of infants, consists of a stimulus generator designed to transmit at least one true random stimulus sequence into a subject's inner ear, a detector for auditory evoked potential (AEP) signals (which yield a set of waveforms), a detector



of evoked otoacoustic emission (OAE) signals (which result in another set of waveforms), and a signal analyzer for assessing the AEP and OAE signals. The signal analyzer incorporates averagers for each of these two sets of waveforms and a synchronizer for coordinating the stimulus generator and signal analyzer.—DRR



vides an opportunity for specific sounds to be elicited from the user. Sounds captured by the computer audio system are analyzed for fairly traditional phonetic patterns, such as the formant frequency and onset time traces as shown in the figure.—DLR

6,435,876

43.70.Kv INTERACTIVE LEARNING OF A FOREIGN LANGUAGE

Timothy Xun Chen, assignor to Intel Corporation
20 August 2002 (Class 434/157); filed 2 January 2001

This foreign language instruction system would involve video recording of the user's mouth movements and video presentation of prerecorded movements by a native speaker as well as recording and playback of the user's and native speaker's pronunciations of difficult sounds. However, the analysis is not done by computer. The system is designed to facilitate interaction over a computer network between the user and a live human language instructor.—DLR

6,435,877

43.70.Ep PHONOLOGICAL AWARENESS, PHONOLOGICAL PROCESSING, AND READING SKILL TRAINING SYSTEM AND METHOD

Janet M. Wasowicz, assignor to Cognitive Concepts, Incorporated
20 August 2002 (Class 434/167); filed 20 July 2001

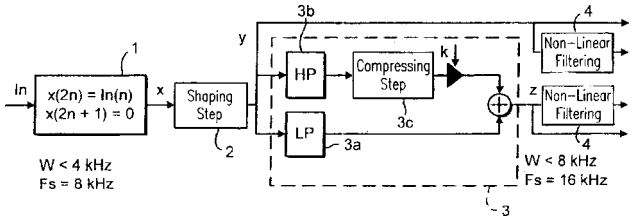
This interactive computer system would allow the clinical assessment and user practicing of phonetic production, particularly for use by individuals who have shown learning difficulties. A video game environment pro-

6,507,820

43.72.Ew SPEECH BAND SAMPLING RATE EXPANSION

Petra Deutgen, assignor to Telefonaktiebolaget LM Ericsson
14 January 2003 (Class 704/500); filed in the United Kingdom 6 July 1999

After interpolating, the resulting spectrum will have aliases up and down. As shown, the inventor proposes the use of a ''shaping filter'' to offer



a frequency weighted attenuation. He also uses a compressor so that unvoiced signals are not unduly filtered.—MK

6,492,913

43.72.Gy METHOD AND CIRCUIT FOR DECODING AN ANALOG AUDIO SIGNAL USING THE BTSC STANDARD

Matthias Vierthaler *et al.*, assignors to Micronas GmbH
10 December 2002 (Class 341/50); filed 17 August 2001

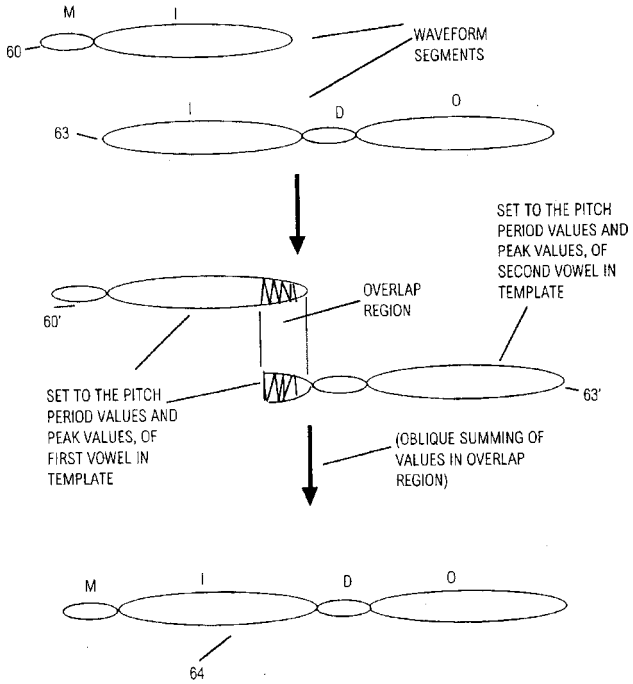
An integrated circuit for decoding an analog audio signal is comprised of a tuner that receives the analog audio signal and provides a sound intercarrier frequency signal. A digital demodulator demodulates the sound intercarrier frequency signal resulting in a digitized multichannel demodulated TV sound signal (MTS). The MTS signal is decoded by a digital broadcast television system committee-compatible (BTSC) decoder, whose audio output consists of summed (L+R) and difference (L−R) signals. Because of this digitally based design method, no adjustments are said to be needed during manufacture or subsequently due to component drift.—DAP

6,438,522

43.72.Ja METHOD AND APPARATUS FOR SPEECH SYNTHESIS WHEREBY WAVEFORM SEGMENTS EXPRESSING RESPECTIVE SYLLABLES OF A SPEECH ITEM ARE MODIFIED IN ACCORDANCE WITH RHYTHM, PITCH AND SPEECH POWER PATTERNS EXPRESSED BY A PROSODIC TEMPLATE

Toshimitsu Minowa *et al.*, assignors to Matsushita Electric Industrial Company, Limited
20 August 2002 (Class 704/258); filed in Japan 30 November 1998

This system for the synthesis of Japanese speech by waveform segment concatenation includes a database of prosodic templates which specify syllable (mora) duration, pitch, and amplitude. A prosodic model generates



the duration, pitch, and amplitude parameters based on mora counts and accent type. As waveform segments are retrieved for the synthesis, they are modified so as to produce the desired prosodic structures.—DLR

6,442,523

43.72.Ja METHOD FOR THE AUDITORY NAVIGATION OF TEXT

Steven H. Siegel, Brooklyn, New York
27 August 2002 (Class 704/270); filed 16 May 2000

This speech synthesis system, intended for children or others perhaps not familiar with the written language, is organized for rapid, interactive selection on a computer screen of the words to be synthesized. This allows a rapid summary-type presentation of screen displays produced by a sequence of mouse movements without clicks. Several variations of the basic system would include an emphasis on phonetic responses and the ability to display graphic symbols under various conditions.—DLR

6,446,040

43.72.Ja INTELLIGENT TEXT-TO-SPEECH SYNTHESIS

Gudrun Socher *et al.*, assignors to Yahoo! Incorporated
3 September 2002 (Class 704/260); filed 17 June 1998

This is a text preprocessor designed to produce input strings for a text-to-speech synthesizer. The text is initially formatted as XML code, with the intention of using the tag mechanism to provide greater flexibility to drive various synthesizers. Semantic information may be attached to one or more words in the text during the original entry or via a database lookup.

User information may be added to control certain synthesis output characteristics. Controllable speech traits include pitch, volume, rate, gender, and emphasis on particular words.—DLR

6,446,038

43.72.Kb METHOD AND SYSTEM FOR OBJECTIVELY EVALUATING SPEECH

Aruna Bayya and Marvin Vis, assignors to Qwest Communications International, Incorporated
3 September 2002 (Class 704/232); filed 1 April 1996

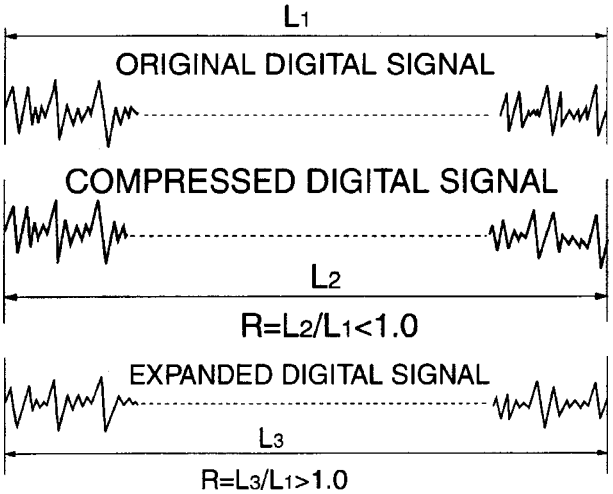
This device uses a neural network to assess speech distortion features extracted from a speech signal affected by a communication system under test. A set of appropriate features is first determined from a clean speech signal. The network is trained by processing speech signals with known defects or by subjecting clean speech to a channel with known distorting characteristics. The goal is to provide an objective measure of the subjective characteristics observed in distorted speech.—DLR

6,487,536

43.72.Lc TIME-AXIS COMPRESSION/EXPANSION METHOD AND APPARATUS FOR MULTICHANNEL SIGNALS

Shinji Koezuka and Kazunobu Kondo, assignors to Yamaha Corporation
26 November 2002 (Class 704/500); filed in Japan 22 June 1999

In time compression systems, because discontinuities can occur in spliced portions of the cut-out waveform segments, cross fading is used to smooth these spliced portions. If the cross-fading is not fully synchronized between left and right channel signals, phase mismatch may cause an unnatural auditory localization for the listener. Disclosed is a method for time compression/expansion for a multichannel system in which waveform segments are cut out sequentially from each channel's signal without causing a lack of synchronization between cross-fading points of the channels. A best cutting starting point for all channels is determined from a synthesized signal formed by averaging the best cutting starting points for all the channel signals.—DAP



6,434,522

43.72.Ne COMBINED QUANTIZED AND CONTINUOUS FEATURE VECTOR HMM APPROACH TO SPEECH RECOGNITION

Eiichi Tsuboka, Neyagawa-shi, Osaka, Japan
13 August 2002 (Class 704/256); filed in Japan 18 June 1992

Discrete probability distribution hidden Markov models (HMMs), which are used to convert input feature vectors to labels, have a computational advantage over continuous probability distribution HMMs, which use input feature vectors as is. However, systems using discrete HMMs have the disadvantage of lower speech recognition accuracy compared to that of continuous HMM systems. Disclosed is a HMM-based speech recognition system said to have improved speech recognition accuracy and a reduced number of calculations. A continuous probability distribution HMM is generated first with a vector quantizing circuit, a set of vectors representing a set of learned patterns is clustered, and a discrete probability distribution HMM is then generated based on the occurrence of the vectors in a cluster.—DAP

6,434,524

43.72.Ne OBJECT INTERACTIVE USER INTERFACE USING SPEECH RECOGNITION AND NATURAL LANGUAGE PROCESSING

Dean Weber, assignor to One Voice Technologies, Incorporated
13 August 2002 (Class 704/257); filed 5 October 1999

This speech recognition system uses customized, context-specific grammars and a natural language processor to improve the recognition performance across a variety of situations. The application is the control of a media access system, including television and Internet capabilities, to obtain news, weather, entertainment, and to perform financial transactions. For each application area, a data definition file specifies particular grammars, vocabularies, and semantic interpretations of recognized words and phrases. In addition, the network context allows considerable flexibility in the implementation.—DLR

6,434,526

43.72.Ne NETWORK APPLICATION SOFTWARE SERVICES CONTAINING A SPEECH RECOGNITION CAPABILITY

Frank Cilurzo and Roger Matthew Miller, assignors to International Business Machines Corporation
13 August 2002 (Class 704/270.1); filed 29 June 1998

This is a fairly typical network-based voice processing system. The user's speech data, entered at a client's terminal, is identified as such and transmitted to a central location where a speech server is provided. In this case, presumably intended for a small-to-moderate sized office, all user speech files, such as voice models and specific vocabulary, are maintained at the central server. As voice processing components, the system would use previously known products of the assignee company.—DLR

6,434,527

43.72.Ne SIGNALLING AND CONTROLLING THE STATUS OF AN AUTOMATIC SPEECH RECOGNITION SYSTEM FOR USE IN HANDSFREE CONVERSATIONAL DIALOGUE

Eric Horvitz, assignor to Microsoft Corporation
13 August 2002 (Class 704/275); filed 17 May 1999

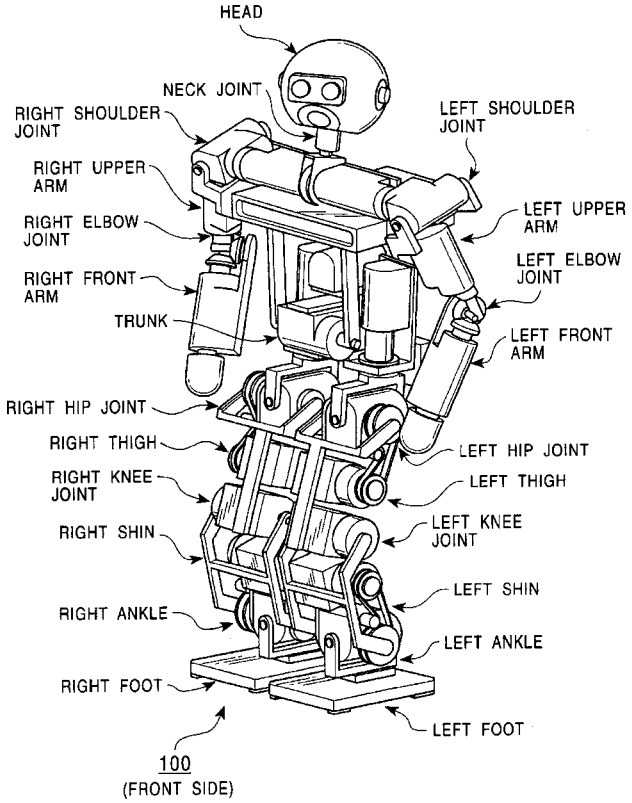
This interactive speech recognition system includes a module which would always remain active and which would determine when user voice input is expected. At such a time, the recognition system would be turned on. However, the general problem of listening at all times for a spoken computer activation command is not addressed. The listening mode described here is a time period following a separate computer action, such as a separately activated process or the computer asking a question to be answered by the user.—DLR

6,438,454

43.72.Ne ROBOT FAILURE DIAGNOSING SYSTEM

Yoshihiro Kuroki, assignor to Sony Corporation
20 August 2002 (Class 700/245); filed in Japan 25 November 1999

Following a long and detailed background discussion of the hopes and goals of those who work in the field of robotics, this patent deals very specifically with the query for and response of diagnostic information. The



goal is to achieve sufficient generality in the interface such that the user may either give a specific diagnosis command or may simply ask, ''How do you feel?'' The robot may reply with specific diagnostic information or may simply say, ''I feel fine.'' or, for example, ''My leg hurts.''—DLR

6,438,523

43.72.Ne PROCESSING HANDWRITTEN AND HAND-DRAWN INPUT AND SPEECH INPUT

John A. Oberteuffer, Lexington, Massachusetts *et al.*
20 August 2002 (Class 704/270); filed 18 May 1999

This patent covers a process for concurrent recognition of written and speech inputs in a PDA-type device. Various modes of operation are described. For example, in an error-correcting context, the same or portions of the same content might appear on both channels. A figure might be sketched and then voice used to describe how the figure is to be modified or enhanced. A signature or drawing might be placed by voice into other artwork. Menu or ''pointing'' commands may be entered using either or both modalities.—DLR

6,438,524

43.72.Ne METHOD AND APPARATUS FOR A VOICE CONTROLLED FOREIGN LANGUAGE TRANSLATION DEVICE

Guangming Carl Shi, assignor to Qualcomm, Incorporated
20 August 2002 (Class 704/277); filed 23 November 1999

This small, hand-held device would accept a spoken phrase in the user's language and, upon recognition of that phrase, would play a suitable translation in a foreign language. Storage is provided for translations in various languages and access is provided for uploading from external storage media. In spite of hints suggesting another mode, clearly the dominant usage mode is a trained, speaker-dependent recognition method. How long will it be before such a device will also listen to the foreign language reply and translate that back into the user's language?—DLR

6,442,524

43.72.Ne ANALYZING INFLECTIONAL MORPHOLOGY IN A SPOKEN LANGUAGE TRANSLATION SYSTEM

Doris M. Ecker *et al.*, assignors to Sony Corporation; Sony Electronics Incorporated
27 August 2002 (Class 704/277); filed 29 January 1999

This spoken language translation system places the emphasis on a detailed morphological analysis of the spoken input prior to the semantic operations required for the translation. The morphological analysis involves a set of features which participate in both the lexical access and the semantic analysis. The underlying linguistic architecture involves F structures and J structures similar to those described in the lexical-functional model.—DLR

6,449,496

43.72.Ne VOICE RECOGNITION USER INTERFACE FOR TELEPHONE HANDSETS

Scott D. Beith *et al.*, assignors to Qualcomm Incorporated
10 September 2002 (Class 455/563); filed 8 February 1999

This application of a collection of well-known voice interface techniques serves to implement a comprehensive voice dialing system for a cell

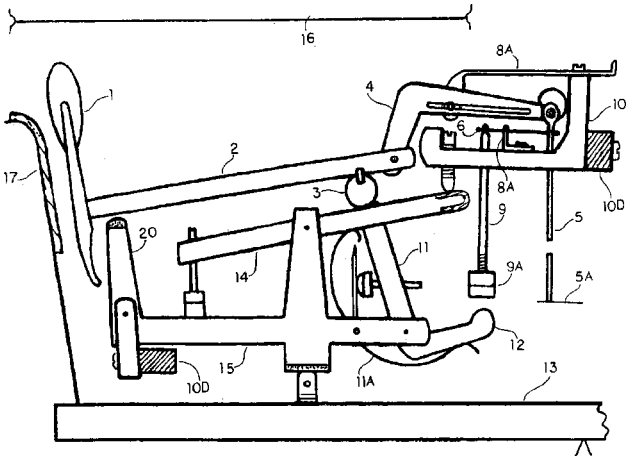
phone or similar portable device. The speaker-dependent recognizer is trained following a protocol using voice, key presses, and prior information. The patent is almost entirely concerned with the menu sequences and includes extensive flow charts of training and usage patterns. The question of adapting to speaker characteristics during use is not addressed. The only issue dealing with recognizer operation, *per se*, is a discussion of how the system would handle interrupted prompts. Rather than using the fairly standard mechanism often called ''barge-in,'' where the prompt is aborted and the entire user utterance is captured, a technique referred to as ''spoke too soon'' is described in which the user's utterance is truncated and an error recovery process begins. Perhaps this indicates a problem with the acoustics of such small device enclosures.—DLR

6,509,517

43.75.Mn GRAND PIANO ACTION

Richard Wroblewski, Trumbull, Connecticut
21 January 2003 (Class 84/216); filed 8 June 2000

The problem at hand is the lack of full dynamic control in a piano. A conventional piano can shift the hammer to not hit as many strings, but then



the touch is not the same. Accordingly, the inventor proposes the use of new action that uses an adjustable fulcrum.—MK

6,506,967

43.75.St PRERECORDED MUSICAL ARRANGEMENT TEACHING ASSEMBLY AND METHOD

Patrick M. Calello, Newark, New Jersey
14 January 2003 (Class 84/470 R); filed 8 June 2000

Four-part *acapella* singing requires good coordination—just ask the Temptations. To assist in training singers in multipart harmonies, the inventor proposes storing multiple versions of a song with each voice being eliminated, one by one. The whole scheme is reminiscent of ''Music Minus One'' except for four voices. But then why not just store each voice separately and then mix them at rehersal time?—MK

6,499,560

43.80.Qf DISPOSABLE STETHOSCOPE DIAPHRAGM AND DISPENSER

Heinrich Lang, The Woodlands and Izaak A. Ulert, Houston, both of Texas
31 December 2002 (Class 181/131); filed 9 June 2000

The apparent purpose of this device is to accommodate disposable diaphragms which can be changed between uses so that only clean and uninfected portions of the stethoscope come in contact with the patient. The patent also covers means for retrofitting existing stethoscopes. The disposable diaphragm assembly may be made of plastic, rubber, or other materials.—DRR

6,503,202

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR FLOW ANALYSIS

John A. Hossack *et al.*, assignors to Acuson Corporation
7 January 2003 (Class 600/454); filed 29 June 2000

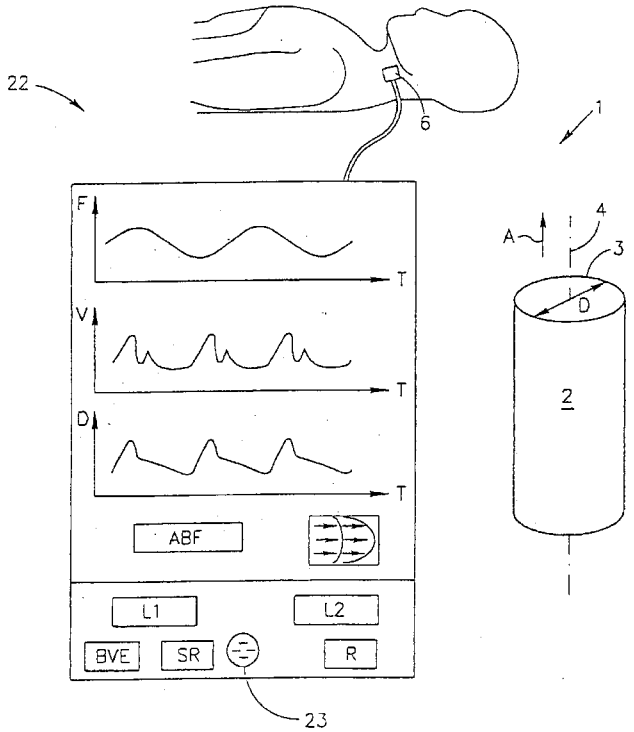
This version of a diagnostic ultrasound system is aimed toward blood flow analysis, particularly in regard to detection of arterial stenosis or constriction in a vessel. Multiple cross-sectional areas along a vessel are determined automatically. A processor locates an abnormality as a function of the multiple cross-sectional areas by determining a cross-sectional area that is a threshold amount less than an average cross-sectional area. The abnormal area is highlighted on the display. Other methods of user notification may also be provided (e.g., an audible warning).—DRR

6,503,205

43.80.Qf DUAL ULTRASONIC TRANSDUCER PROBE FOR BLOOD FLOW MEASUREMENT, AND BLOOD VESSEL DIAMETER DETERMINATION METHOD

Dan Manor *et al.*, assignors to Cardiosonix Limited
7 January 2003 (Class 600/459); filed in Israel 18 November 1998

This device constitutes a dual ultrasonic transducer probe for use in a Doppler-based ultrasound system for blood flow measurement, determination of associated hemodynamic parameters, and measurement of the blood vessel diameter. Two transducers in a probe are oriented with respect to each



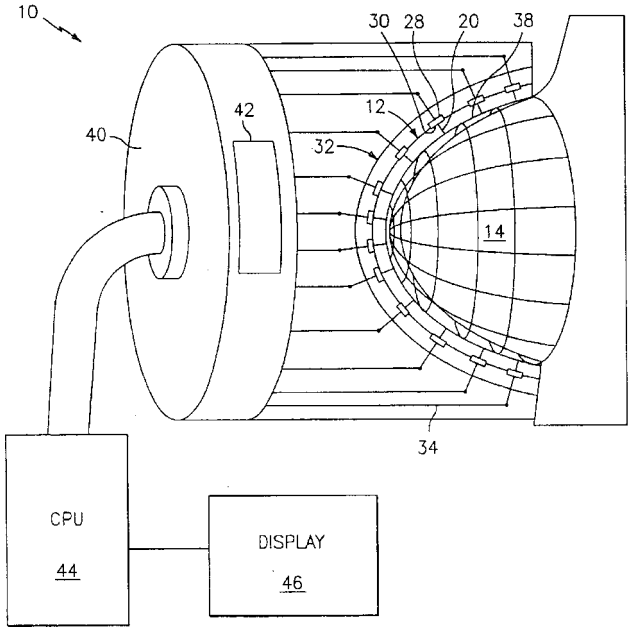
other so that the ultrasound beams intercept at a certain acute angle and at the blood vessel's longitudinal axis. Determination of the blood vessel diameter is based on the creation of a two-dimensional matrix of complex reflection amplitude values for each of the beams along its axis of propagation.—DRR

6,511,433

43.80.Qf ACTIVE ACOUSTIC ARRAY FOR ULTRASONIC BIOMEDICAL APPLICATIONS

Kim C. Benjamin, assignor to the United States of America as represented by the Secretary of the Navy
28 January 2003 (Class 600/463); filed 30 July 2001

The purpose of this acoustic array is to execute biomedical assessments, particularly those for detecting cancer in human tissue. The device consists of an array to conform to and surround a portion of the human anatomy and a material for acoustically coupling the array to the human anatomy portion. The array itself consists of multiple doubly curved segments, each segment formed by a piezoelectric ceramic polymer composite material with an acoustic element formed on one surface by selective deposition of a conductive material. The acoustic elements in the pattern function as both transmitters and receivers. The acoustic array includes a backing material to provide mechanical damping and to define the shape of the array. The device also contains signal conditioning electronics for processing signals received from the array. A central processing unit is provided to generate cross-sectional images of the human tissues under examination, which may be displayed on video.—DRR



6,512,830

43.80.Qf AUSCULTATION APPARATUS

Birger Orten, assignor to Meditron AS
28 January 2003 (Class 381/67); filed in Norway 9 February 1995

An auscultation device is an apparatus to listen to sounds from internal regions of a living body, principally the heart, respiratory, or circulatory organs and, in this case, to analyze the sounds in a more detailed fashion for diagnostic purposes. Essentially the device is an electronic stethoscope that includes a microphone, amplifier, loudspeaker, and an adjustable bandpass filter circuit. The user can adjust the center frequency and bandwidth of the filter at will throughout the audible range.—DRR

6,500,121

43.80.Sh IMAGING, THERAPY, AND TEMPERATURE MONITORING ULTRASONIC SYSTEM

Michael H. Slayton and Peter G. Barthe, assignors to Guided Therapy Systems, Incorporated
31 December 2002 (Class 600/439); filed 10 February 2000

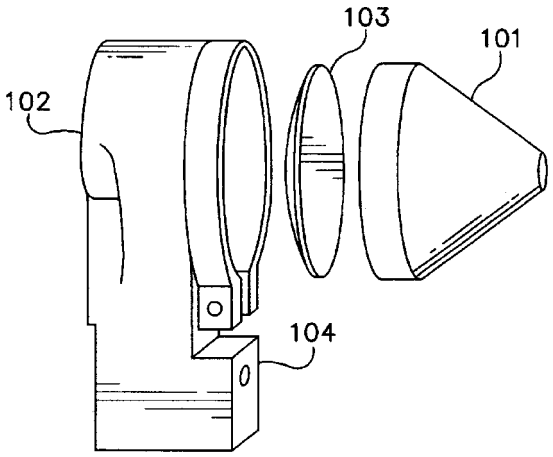
This relates to a noninvasive therapeutic ultrasonic system that is capable of acoustically imaging and heating a certain region to be treated and also of acoustically monitoring the temperature profile of the treatment region. The acoustic transducer assembly consists of a single transducer in an imaging subsystem, a therapy subsystem, and a temperature monitoring subsystem. The ultrasound system may include a display for imaging and temperature monitoring functions. The acoustic transducer assembly may also be configured to provide three-dimensional imaging, temperature monitoring, or therapeutic heating through the use of adaptive algorithms and/or rotational or translational movement. Treatment capabilities may be further extended through the use of additional single transducers.—DRR

6,500,133

43.80.Sh APPARATUS AND METHOD FOR PRODUCING HIGH INTENSITY FOCUSED ULTRASONIC ENERGY FOR MEDICAL APPLICATIONS

Roy W. Martin *et al.*, assignors to University of Washington
31 December 2002 (Class 601/3); filed 1 December 2000

This medical instrument makes use of solid or tapered cones mounted to a concave, spherically curved piezoelectric ultrasonic transducer, which



focuses and concentrates the ultrasonic energy into a narrow tip so that high intensities of ultrasound can be delivered to the tissue adjacent to the tip.—DRR

6,500,141

43.80.Sh APPARATUS AND METHOD FOR TREATING BODY TISSUE, IN PARTICULAR SOFT SURFACE TISSUE WITH ULTRASOUND

Klaus M. Irion and Nicanor G. Isse, assignors to Karl Storz GmbH & Company KG
31 December 2002 (Class 604/22); filed in Germany 8 January 1998

This is another device for applying ultrasound for treating body tissues. The device appears to differ principally by featuring a suction cup that sucks the body surface against the ultrasound applicator surface.—DRR

6,506,157

43.80.Sh DUAL DOPPLER ARTERY LIGATION AND HEMORRHOID TREATMENT SYSTEM

Jack Teigman and John M. D'Angelo, both of St. Petersburg, Florida
14 January 2003 (Class 600/439); filed 5 September 2001

This dual Doppler artery ligation and hemorrhoid treatment system is equipped with a probe that has a disposable cylinder with a cone-shaped cap at the distal end and a support adjacent to the proximal end. A series of white light emitting diodes is mounted within the support. The support contains a proximal backing plate and a distal opening to the cylinder and cone. The cylinder has an opening next to its distal end and a pair of ultrasonic transducers attached to the cylinder laterally spaced on the centerline at the proximal side of the opening. An echo analyzer connected to the probe is configured to emit signals and to detect and analyze received signals from the blood flow. The detected signals produce an output that varies as a function of the arterial blood flow of a hemorrhoid.—DRR

6,506,171

43.80.Sh SYSTEM AND METHODS FOR CONTROLLING DISTRIBUTION OF ACOUSTIC ENERGY AROUND A FOCAL POINT USING A FOCUSED ULTRASOUND SYSTEM

Shuki Vitek and Naama Brenner, assignors to InSightec-TxSonics, Limited
14 January 2003 (Class 601/2); filed 27 July 2000

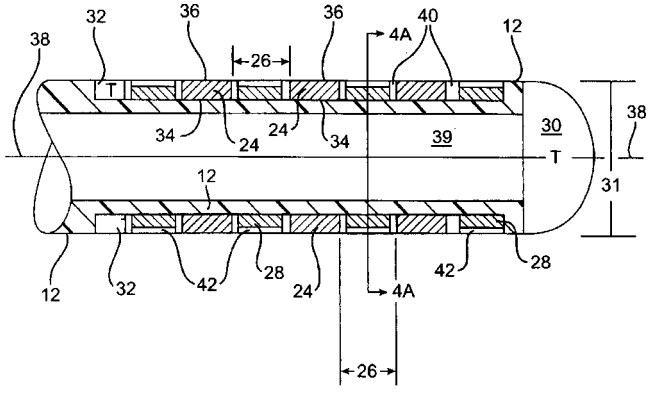
This is a system for performing noninvasive surgical procedures through the use of ultrasound and for controlling distribution of acoustic energy in the vicinity of a focal point, e.g., to create uniform necrosis volume around the focal point via a focused ultrasound transducer array. Multiple transducer elements are disposed around a central axis. Drive circuits are coupled to the transducer elements. A controller, coupled to the drive circuitry, programs the drive circuitry to drive the transducers in such a manner to have the appropriate phase shift values based on each transducer element's angular position.—DRR

6,508,765

43.80.Sh ULTRASOUND-GUIDED ABLATION CATHETER AND METHODS OF USE

Veijo T. Suorsa *et al.*, assignors to SCIMED Life Systems, Incorporated
21 January 2003 (Class 600/439); filed 28 December 2000

In one embodiment of this ultrasound-guided ablation catheter, the tissue ablation apparatus includes a flexible elongated body. Drive signals are applied to one set of transducers to emit therapeutic ultrasonic waves focused at a focal location, thereby heating tissues at that location. The process also includes acquiring received signals generated by a second set of transducers. The received signals can be summed or otherwise processed to



yield a detected signal representative of the ultrasonic waves reflected or emanating from the focal location. A feedback signal is also produced based on the detected signal indicating that cavitation has occurred or is likely to occur within the subject volume in the area of the focal location. The emitted therapeutic ultrasonic bombardment waves are altered or terminated in response to the feedback signal.—DRR

6,508,774

43.80.Sh HIFU APPLICATIONS WITH FEEDBACK CONTROL

David E. Acker and Patrick David Lopath, assignors to Transurgical, Incorporated
21 January 2003 (Class 601/2); filed 9 March 2000

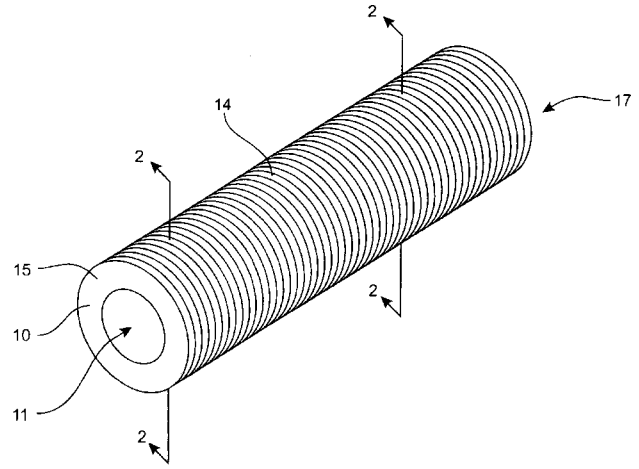
An array of transducers is used to apply therapeutic ultrasonic waves to produce high-intensity focused ultrasound (HIFU) heating. Cavitation is detected by using some or all of the transducers in the array to detect ultrasonic waves emanating from or reflected from the patient's body or the interface. A feedback signal is generated based on such detection. If the feedback signal indicates the occurrence of cavitation, the therapeutic ultrasonic waves are either terminated or adjusted. The same transducers can also detect obstacles to sound transmission, such as bony structures or bubbles at the interface between the transducer array and the patient.—DRR

6,508,775

43.80.Sh HIGH OUTPUT THERAPEUTIC ULTRASOUND TRANSDUCER

John McKenzie and Joseph Karratt, assignors to Pharmasonics, Incorporated
21 January 2003 (Class 601/2); filed 20 March 2001

The subject therapeutic ultrasound delivery system consists of a catheter body and a plurality of axially spaced, hollow, cylindrical, vibrational piezoelectric transducers distributed along the length of the catheter body. A spring restraint wrapped around the outer surfaces of the transducers exerts



an inward prestress on the outer surfaces of these transducers. Another spring restraint is in contact with the inner surfaces of the transducers. It is claimed that this transducer configuration can yield very high amplitudes without the transducer elements breaking apart.—DRR

6,508,781

43.80.Sh ULTRASONIC ABLATION CATHETER TRANSMISSION WIRE CONNECTOR ASSEMBLY

Lawrence Brennan *et al.*, assignors to Advanced Cardiovascular Systems, Incorporated
21 January 2003 (Class 604/22); filed 30 December 1999

This ultrasonic angioplasty device consists of a catheter and a connector assembly for interconnecting an ultrasonic transmission member to an ultrasound transducer. A unique feature of the device is a rotation control that prevents the ultrasound transmission member from rotating relative to the catheter. Rotation of the connector assembly relative to the catheter is prevented by a fastener aligning the connector assembly with the elongated catheter body or by a key that fits into a correspondingly shaped slot in a proximal shaft portion of the catheter.—DRR

6,508,783

43.80.Sh ULTRASOUND METHOD FOR REVASCULARIZATION AND DRUG DELIVERY

Lauri J. DeVore, assignor to Scimed Life Systems, Incorporated
21 January 2003 (Class 604/22); filed 14 March 2001

This is a method for generating an angiogenic response in an interior wall of the heart using a percutaneous myocardial revascularization (PMR) procedure. The patent describes an intravascular ultrasound PMR device having a needle for supplying the heart with an angiogenic material or contrasting agent. The ultrasound energy is applied to the area in need of improved circulation. A distal head on an elongated tubular body introduces ultrasound to the region of interest and a needle delivers angiogenic materials to an area of interest.—DRR

6,511,432

43.80.Sh PREAMPLIFIER AND PROTECTION CIRCUIT FOR AN ULTRASOUND CATHETER

Thomas C. Moore *et al.*, assignors to Scimed Life Systems, Incorporated
28 January 2003 (Class 600/459); filed 27 April 2001

The device consists of an ultrasound catheter featuring a rotatable transducer distally connected to a transmission line. A preamplifier is connected to the transmission line and coupled to the rotatable transducer. Protection circuitry protects the input and output of the preamplifier from high-amplitude electrical impulses resulting from the transducer excitation signal. The protection circuitry may consist of either back-to-back diodes or diode bridges, or it may comprise at least one unit of two GaAs switches coupled to the input and output of the preamplifier, respectively.—DRR

6,511,444

43.80.Sh TRANSMYOCARDIAL REVASCULARIZATION USING ULTRASOUND

Kullervo Hynynen *et al.*, assignors to Brigham and Women's Hospital
28 January 2003 (Class 601/2); filed 31 January 2000

A method and apparatus are disclosed for executing transmyocardial revascularization. A phased array ultrasonic device consists of multiple ultrasonic transducer elements that are controlled using a feedback control

system so that each ultrasonic transducer element produces an ultrasonic wave of specific power and phase in order to achieve constructive interference at a desired acoustic focus. The constructive interference generates high-pressure amplitudes causing a rapid rise in tissue temperature which vaporizes the target tissue at the focus.—DRR

6,494,836

43.80.Vj METHOD OF GENERATING IMAGE DATA AND ULTRASONIC DIAGNOSTIC APPARATUS USING THE SAME

Eiji Ogawa, assignor to Fuji Photo Film Company, Limited
17 December 2002 (Class 600/443); filed in Japan 15 May 2000

Three-dimensional image data is obtained in three steps. First, signals from sensors are stored. Second, image data is obtained around a measured point in the object to be imaged. Third, steps one and two are repeated for different points in the object.—RCW

6,494,837

43.80.Vj SYSTEM AND METHOD FOR THREE-DIMENSIONAL ULTRASOUND IMAGING USING A STEERABLE PROBE

Sang-Hyun Kim and Seok-Bin Ko, assignors to Medison Company, Limited
17 December 2002 (Class 600/443); filed in the Republic of Korea 10 June 2000

A three-dimensional image is produced by quickly computing the distance between adjacent cross-sectional images. In the processing, two parallel frames and a supplementary frame inclined at an angle with respect to the parallel frames are used. Correlation coefficients between frames are computed to estimate the distance between the two parallel frames.—RCW

6,500,125

43.80.Vj ULTRASOUND B/COLOR PRIORITY THRESHOLD CALCULATION

David J. Muzilla and Gary E. MacLeod, assignors to GE Medical Systems Global Technology Company, LLC
31 December 2002 (Class 600/454); filed 30 March 2000

This calculation uses b-mode data in the region of valid color data to determine a threshold for combined display of color flow and b-mode information with reduced color bleeding artifacts.—RCW

6,500,134

43.80.Vj METHOD FOR TREATING CIRCULATORY DISORDERS WITH ACOUSTIC WAVES

Alphonse Cassone, Las Vegas, Nevada
31 December 2002 (Class 601/47); filed 19 July 2000

This method for treating circulatory diseases consists of operating a low-frequency acoustic transducer in a liquid-containing vat, positioning the patient suffering a circulatory disorder at a ''therapeutically beneficial'' distance from the transducer and exposing that person for a ''therapeutically beneficial'' period of time at some unspecified low frequency in order to treat inflammatory musculoskeletal connective tissue disorders. In a more specific embodiment, the patient is positioned between 1 and 20 ft from the low-frequency transducer for an exposure of between 15 and 25 min at a frequency of approximately 600 Hz. Whether this device works, given the attenuation that must occur in the emanation of the sound waves and the audio frequencies of the applied sound, is open to question.—DRR

6,503,199

43.80.Vj UNIFORM VOLUMETRIC SCANNING ULTRASONIC DIAGNOSTIC IMAGING SYSTEM

Daniel Lennon, assignor to ATL Ultrasound
7 January 2003 (Class 600/443); filed 3 November 1999

An image plane of steered scanning beams is rotated around an axis through a volume with beams in the vicinity of the axis more widely separated than beams farther from the axis to obtain a more uniform sampling density throughout the scanned volume.—RCW

6,503,203

43.80.Vj AUTOMATED ULTRASOUND SYSTEM FOR PERFORMING IMAGING STUDIES UTILIZING ULTRASOUND CONTRAST AGENTS

Patrick G. Rafter and Heinrich Beckermann, assignors to Koninklijke Philips Electronics N.V.
7 January 2003 (Class 600/458); filed 16 January 2001

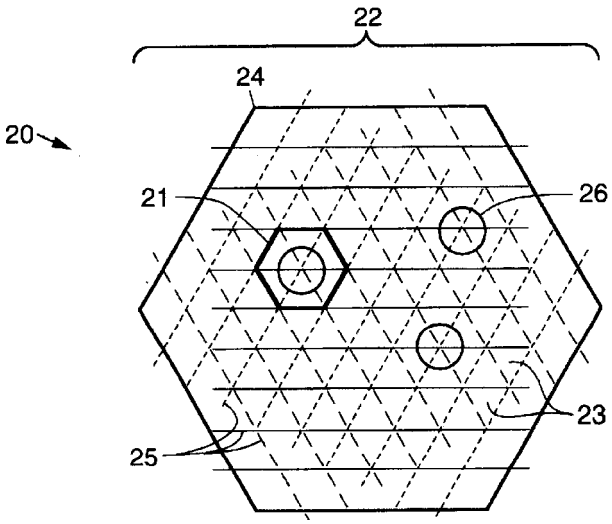
Control circuitry sequentially adjusts imaging parameters through a sequence of values during an ultrasound imaging study. The parameter sequence is accessible by the control circuitry to conduct other imaging studies. This facilitates rapid alteration of imaging parameters during ultrasonic studies, for example, with and without contrast.—RCW

6,503,204

43.80.Vj TWO-DIMENSIONAL ULTRASONIC TRANSDUCER ARRAY HAVING TRANSDUCER ELEMENTS IN A NON-RECTANGULAR OR HEXAGONAL GRID FOR MEDICAL DIAGNOSTIC ULTRASONIC IMAGING AND ULTRASOUND IMAGING SYSTEM USING SAME

Thilaka S. Sumanaweera and William Dreschel, assignors to Acuson Corporation
7 January 2003 (Class 600/459); filed 31 March 2000

The elements in this transducer array are arranged in a nonrectangular or hexagonal pattern. The array may be fabricated in a single or multiple



layer structure and sparse arrays may be made by connecting selected transducer elements to an imaging system. In addition, the array may use a random, vernier, or spiral element configuration in a hexagonal grid.—RCW

6,504,288

43.80.Vj COMPENSATED INDIVIDUALLY ADDRESSABLE ARRAY TECHNOLOGY FOR HUMAN BREAST IMAGING

D. Kent Lewis, assignor to The Regents of the University of California
7 January 2003 (Class 310/334); filed 5 December 2000

This technology uses individually addressable array elements for flexible beam formation, digital-to-analog converters for source signals, inverse filtering for source compensation, and spectral extrapolation to expand received signal bandwidth.—RCW

6,505,063

43.80.Vj DIAGNOSTIC IMAGING SYSTEM WITH ULTRASOUND PROBE

Johan Samuel Van Den Brink *et al.*, assignors to Koninklijke Philips Electronics N.V.
7 January 2003 (Class 600/411); filed in the European Patent Office 15 December 1999

An ultrasound probe is included with a magnetic resonance (MR) imaging system. Images reconstructed from MR signals and ultrasound echoes are registered in a common reference frame and geometric distortion in the ultrasonic image is corrected using the MR image. The corrections use different tissue types determined from the MR image to compensate for differences between sound speed in these tissues.—RCW

6,506,156

43.80.Vj ECHOGENIC COATING

Michael L. Jones *et al.*, assignors to Vascular Control Systems, Incorporated
14 January 2003 (Class 600/439); filed 19 January 2000

This coating consists of an insulating base layer and an echogenic layer that includes a polymeric matrix with either voids, glass microspheres, or both.—RCW

6,508,763

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD

Joseph A. Urbano *et al.*, assignors to Ecton, Incorporated
21 January 2003 (Class 600/437); filed 14 September 1999

Field programmable gate arrays are used to reconfigure an ultrasonic imaging system for changes in scanning (e.g., from a phased array to a curved array), for system operating mode changes (e.g., from b-mode to color Doppler), or dynamically as part of system operation (e.g., between transmitted pulses or image frames).—RCW

6,508,764

43.80.Vj ABERRATION CORRECTION APPARATUS AND METHODS

Karl E. Thiele and Bernard J. Savord, assignors to Koninklijke Philips Electronics N.V.
21 January 2003 (Class 600/437); filed 31 October 2000

Transducer array elements with a fine pitch are used with a hierarchial control scheme. The fine pitch of elements or groups of elements is less than or equal to the wavelength at the fundamental frequency and improves aberration correction. Array elements may be placed in subgroups with aberration correction being applied to the output of each subgroup.—RCW

6,508,766

43.80.Vj ULTRASOUND DIAGNOSTIC APPARATUS

Takeshi Sato *et al.*, assignors to Kabushiki Kaisha Toshiba
21 January 2003 (Class 600/441); filed in Japan 20 January 2000

Processors in this apparatus produce b-mode image data and power Doppler data from echo signals. The power Doppler data come from signals acquired under transmission conditions with the number of cycles in the transmit pulse essentially equal to that for the b-mode data. The number of cycles is selectable. The b-mode image data and power Doppler data are partially synthesized.—RCW

6,508,767

43.80.Vj ULTRASONIC HARMONIC IMAGE SEGMENTATION

Peter N. Burns and David Hope Simpson, assignors to Koninklijke Philips Electronics N.V.
21 January 2003 (Class 600/443); filed 12 February 2001

Received ultrasonic echo signals are separated into linear and nonlinear components by a Doppler filter using pulse inversion. A threshold is determined from an estimate of the contributions of linear scattering and noise to the echo. Signals that exceed the threshold are segmented as nonlinear, e.g., originating from microbubbles, and displayed, while signals below the threshold are suppressed.—RCW

6,508,768

43.80.Vj ULTRASONIC ELASTICITY IMAGING

Timothy J. Hall and Yanning Zhu, assignors to University of Kansas Medical Center
21 January 2003 (Class 600/443); filed 17 September 2001

Displacement vectors are estimated from a pattern of samples throughout an imaged region of interest by comparing two successive b-mode frames. From the displacement vectors, corresponding strain values are estimated to describe elasticity of the tissue. An image is displayed showing the strain distribution within the region of interest as it is stressed, for example, by the user pressing the ultrasound transducer against the patient's body.—RCW

6,511,426

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR VERSATILE PROCESSING

John A. Hossack *et al.*, assignors to Acuson Corporation
28 January 2003 (Class 600/437); filed 8 June 1999

This system reduces speckle in cross-sectional and three-dimensional imaging. In two-dimensional imaging, sequential, parallel, or frames at different elevations are compounded. In three-dimensional imaging, two-dimensional frames spaced in elevation are compounded and the frames are used to obtain a three-dimensional set of data, e.g., by interpolation. Also in three-dimensional imaging, Doppler and b-mode data may be combined using different relationships.—RCW

6,512,943

43.80.Vj COMBINED ULTRASOUND-RADIONUCLIDE DEVICE FOR PERCUTANEOUS ULTRASOUND-GUIDED BIOPSY AND METHOD OF USE

Frederick Kelcz, assignor to Wisconsin Alumni Research Foundation
28 January 2003 (Class 600/436); filed 22 May 2000

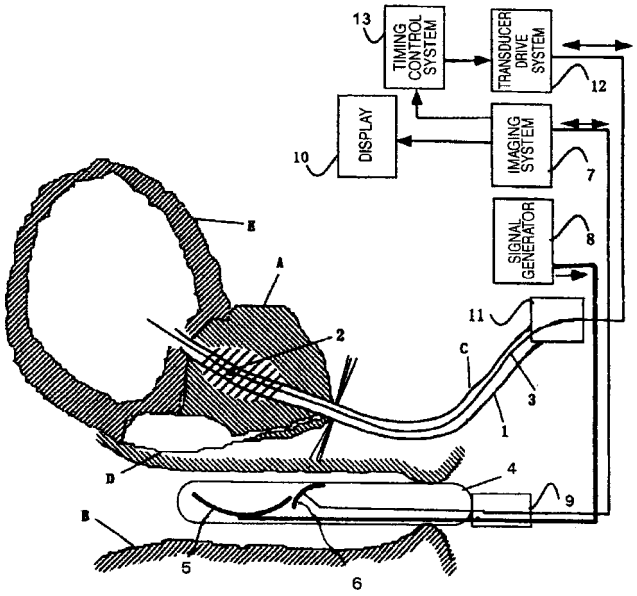
Superimposed, simultaneous ultrasound and nuclear activity images are obtained to help perform an accurate biopsy for procedures in which nuclear activity localizes abnormal tissue. Radionuclide detectors are angled to determine the depth of maximum tracer uptake. This depth is determined by the intersection of lines of sight from the radionuclide detectors. After the depth is determined, an ultrasound image containing the uptake site is produced.—RCW

6,511,428

43.80.Vj ULTRASONIC MEDICAL TREATING DEVICE

Takashi Azuma *et al.*, assignors to Hitachi, Limited; Hitachi Medical Corporation
28 January 2003 (Class 600/439); filed in Japan 26 October 1998

A therapeutic ultrasound beam from a transducer inside a probe is focused in a region to be treated. Another ultrasound transducer inside the probe is used for imaging. The imaging probe is used to show the target



region. The therapeautic transducer is driven in a short-pulse mode at the same frequency as the imaging ultrasound so that a pointlike source is produced and displayed in the image on the target.—RCW

6,511,429

43.80.Vj ULTRASONIC METHODS AND SYSTEMS FOR REDUCING FETAL STIMULATION

Mostafa Fatemi and James F. Greenleaf, assignors to Mayo Foundation for Medical Education and Research
28 January 2003 (Class 600/443); filed 17 August 2000

Audible sound produced by an ultrasonic imaging system is reduced by changing the system operating mode such as by lowering ultrasonic peak power and changing the pulse repetition rate to be outside the audible range.—RCW

6,511,430

43.80.Vj USE OF HIGH FREQUENCY ULTRASOUND IMAGING TO DETECT AND MONITOR THE PROCESS OF APOPTOSIS IN LIVING TISSUES, EX-VIVO TISSUES AND CELL-CULTURE

Michael D. Sherar *et al.*, assignors to University Health Network
28 January 2003 (Class 600/443); filed 19 August 1998

This use of high-frequency ultrasound includes six steps: (1) Imaging a selected site to obtain a ''before'' image. (2) Exposing the selected site to an apoptosis (disintegration into membrane-bound particles)—inducing stress. (3) Imaging the selected site at subsequent time intervals to obtain ''after'' images. (4) Measuring the signal amplitude in a region of the selected site in the before and after images. (5) Comparing the signal amplitude measurements to determine whether the after image exhibits an increase in amplitude relative to the before image to obtain an indication that apoptosis has begun. (6) Measuring the change in a echo spectrum in the region of interest in the before and after images to confirm by an increased slope of the spectrum that apoptosis has begun.—RCW

6,511,427

43.80.Vj SYSTEM AND METHOD FOR ASSESSING BODY-TISSUE PROPERTIES USING A MEDICAL ULTRASOUND TRANSDUCER PROBE WITH A BODY-TISSUE PARAMETER MEASUREMENT MECHANISM

John W. Sliwa, Jr. and Vaughn R. Marian, assignors to Acuson Corporation
28 January 2003 (Class 600/438); filed 10 March 2000

A probe contains a transducer to produce an ultrasonic image. The probe also contains a parameter measurement mechanism such as a mechanically or electrically operated unit for the application of static or dynamic loading on the surface of tissue or at some depth. In addition, a secondary tracking transducer may be included in the probe.—RCW

Scattering of acoustic waves from immersed transversely isotropic cylinders (L)

Farhang Honarvar[a)]
Department of Mechanical Engineering, K. N. Toosi University of Technology, P.O. Box 16765-3881, Tehran, Iran

Ying Fan and Anthony N. Sinclair
Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ontario M5S 3G8, Canada

(Received 5 February 2003; revised 11 April 2003; accepted 25 April 2003)

A mathematical model for the scattering of acoustic waves from immersed transversely isotropic cylinders was developed by the authors [J. Acoust. Soc. Am. **100**, 57 (1996)]. Ahmad and Rahman further elaborated on the mathematical aspects of this model and based on mathematical discussions showed that there exist two distinct types of transversely isotropic materials, which they named type I and type II [Int. J. Eng. Sci. **38**, 325 (2000)]. They also described the expected behavior of the form functions of such materials. The purpose of this article is to distinguish the physical differences between these two types of transversely isotropic materials. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582445]

PACS numbers: 43.20.Fn, 43.20.Ks [ANN]

I. INTRODUCTION

The first mathematical model for acoustic wave scattering from an anisotropic cylinder was developed by the authors in 1996.[1]

Recently, based on mathematical arguments, Ahmed and Rahman showed that, in general, there exist two types (type I and type II) of transversely isotropic materials, which exhibit different scattering characteristics.[2] Based on a mathematical discussion, Ahmad and Rahman also showed that an extraneous critical angle exists in the case of type II materials. No physical interpretation was presented for this extraneous critical angle.

In this article, we describe the reasons for the existence of two distinct types of transversely isotropic materials and use some geometrical reasoning based on slowness curves to provide further evidence for the existence of an extra critical angle in the case of type II materials.

II. CLASSIFICATION OF TRANSVERSELY ISOTROPIC MATERIALS

A transversely isotropic material is characterized by five independent elastic constants, viz., c_{11}, c_{12}, c_{13}, c_{33}, and c_{44}. Such a material is isotropic in one specific plane (say $r-\theta$ plane in a cylindrical coordinate system) and has different elastic properties along the third direction (z axis), i.e., the direction normal to the isotropic plane.[3] It is obvious that the normal stiffness along this direction can be either stronger or weaker than the normal stiffness in the isotropic plane. In what Ahmed and Rahman refer to as type I material, the normal stiffness along the cylinder axis (z-axis) is stronger than the normal stiffness in the isotropic plane and in type II materials, it is the reverse. This difference of the two types of transversely isotropic materials is not explicitly stated in Ref. 2; instead, the existence of an extra critical angle is used as the main sign of difference between these materials.

To further elaborate on this difference, we plot the slowness curves for a number of transversely isotropic materials. The slowness curves for isotropic (aluminum), type I transversely isotropic (cobalt and magnesium), and type II transversely isotropic (titanium boride and zinc) materials are shown in Fig. 1. For isotropic aluminum, all three slowness curves are circles, one corresponding to the longitudinal wave and the other two, which are identical, corresponding to shear waves.[4]

By examining the ratio c_{33}/c_{11} for type I and type II materials in Table 1 of Ref. 2, one notes that for type I materials $c_{33}/c_{11}>1$ and for type II materials $c_{33}/c_{11}<1$. In Ref. 1, it was shown that the stiffness of a transversely isotropic cylinder along the z axis is governed by the two elastic constants c_{33} and c_{44}. The elastic constant c_{33} has a more significant role in controlling the axial stiffness of the cylinder. Similar to the approach followed in Ref. 1, we investigate the difference between type I and type II transversely isotropic materials by perturbing the elastic constants of an isotropic material. Considering isotropic aluminum with Lamé constants $\lambda=61.15$ GPa and $\mu=24.86$ GPa, the five elastic constants corresponding to a transversely isotropic material having the same elastic properties as aluminum are

$$\begin{aligned} c_{11}&=c_{33}=\lambda+2\mu=110.87 \text{ GPa},\\ c_{12}&=c_{13}=\lambda=61.15 \text{ GPa},\\ c_{44}&=\mu=24.86 \text{ GPa}. \end{aligned} \tag{1}$$

Since c_{33} is a crucial parameter in determining whether a transversely isotropic material is of type I or II, this elastic constant is perturbed first by $+30\%$ and then by -30%. The resulting slowness curves are plotted in Figs. 2(b) and (c), respectively. Comparing Figs. 2 and 1, we observe that increasing/decreasing the stiffness of aluminum along the z

a)Electronic mail: honarvar@mie.utoronto.ca

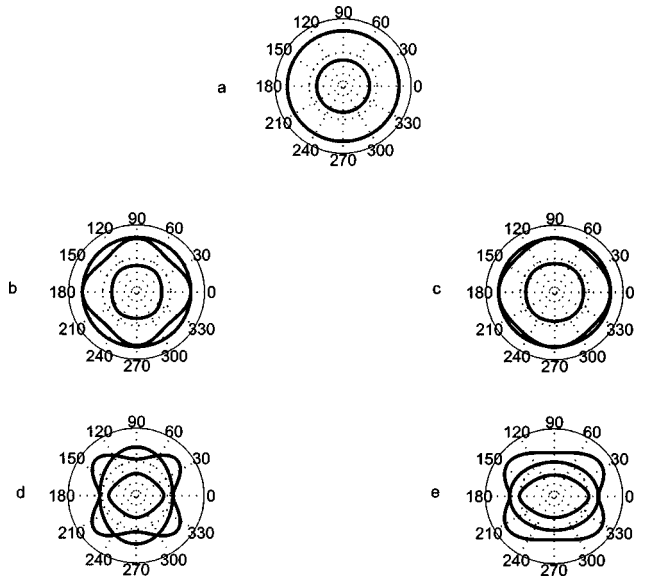


FIG. 1. Slowness curves for (a) aluminum (isotropic), (b) cobalt (type I), (c) magnesium (type I), (d) titanium-boride (type II), and (e) zinc (type II).

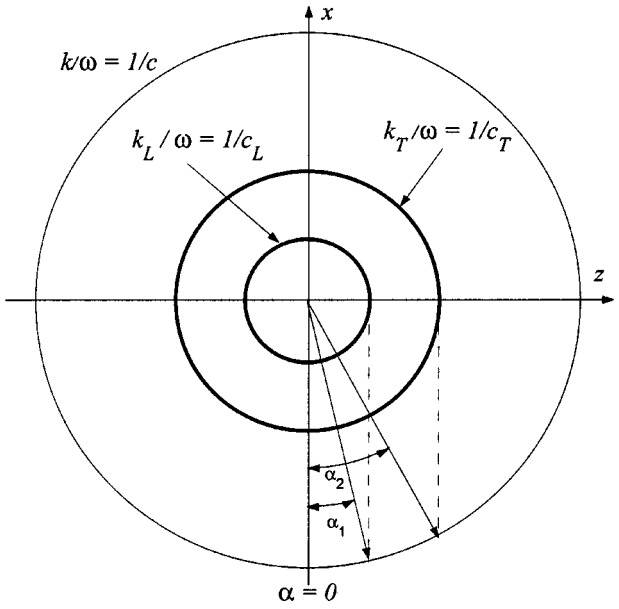


FIG. 3. Geometrical construction for finding critical angles of aluminum.

axis has resulted in slowness curves similar to those of a type I/type II material. Perturbations in various elastic constants of aluminum showed that the ratio c_{33}/c_{11} distinguishes a type I from a type II transversely isotropic material.

III. CRITICAL ANGLES

According to Ahmed and Rahman,[2] for type II transversely isotropic materials, there exists a critical angle beyond which all roots of the characteristic equation are complex. In this section, we will show that similar to the first and second critical angles, this angle can also be verified from the slowness curves of the material.

When studying the transmission of waves from one medium into another, the critical angles can be obtained from slowness curves of the two media.[4] In an isotropic medium, the slowness curve corresponding to the longitudinal or shear wave is a circle with radius equal to the inverse of the phase velocity. The geometrical construction yielding the two critical angles of waves incident from water on aluminum is shown in Fig. 3, where c is the velocity of wave in water and c_L and c_T are the longitudinal and shear wave velocities in aluminum, respectively.

At the boundary of the two media, both fields should have the same component of the wavenumber k tangent to

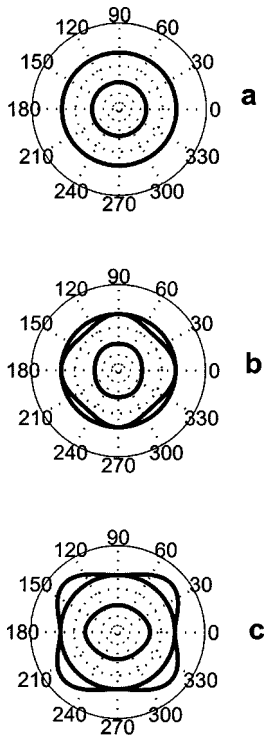


FIG. 2. Slowness curves for (a) aluminum, (b) aluminum with c_{33} perturbed by $+30\%$, and (c) aluminum with c_{33} perturbed by -30%.

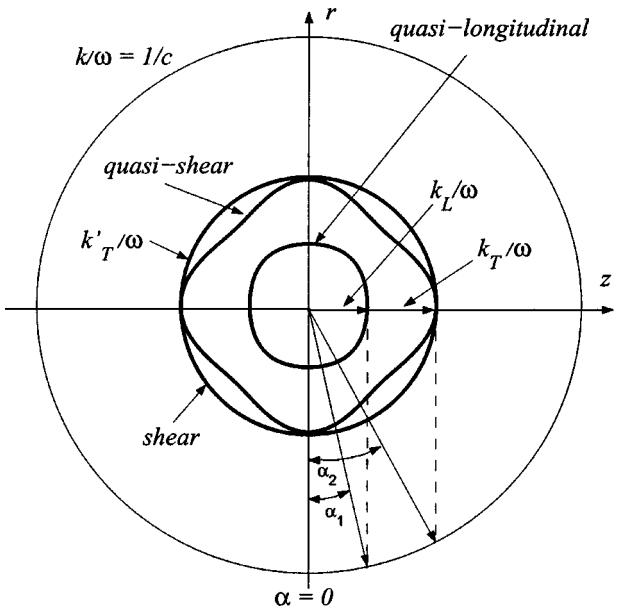


FIG. 4. Geometrical construction for finding critical angles of cobalt (type I transversely isotropic material).

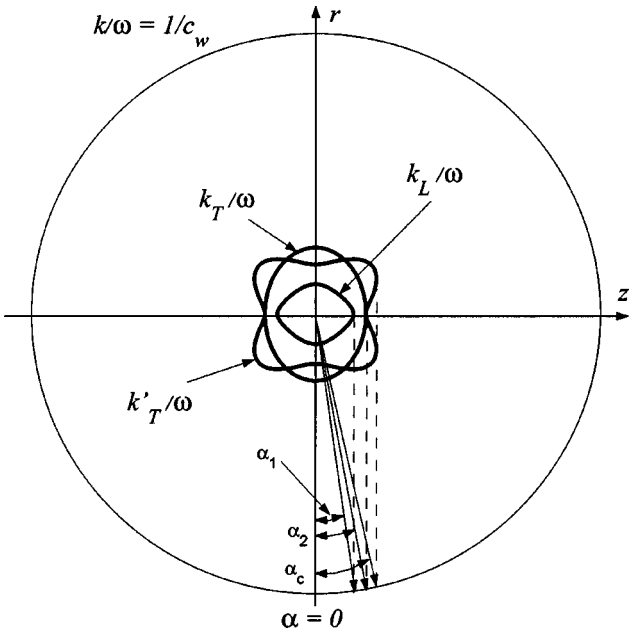


FIG. 5. Geometrical construction for finding critical angles of titanium-boride (type II transversely isotropic material).

the boundary. This component is designated as k_z where z is the coordinate axis parallel to the interface. Therefore, as the incidence angle increases from $\alpha=0$, k_z eventually becomes equal to the radius of the longitudinal/shear wave slowness circle. The incidence angle corresponding to this situation is the first/second critical angle designated by α_1 or α_2. It is noted that at a critical angle the tangent line of the slowness curve is vertical.

Figures 4 and 5 show this geometrical construction applied to type I (cobalt) and type II (titanium-boride) transversely isotropic materials. For type II transversely isotropic materials, there exist a bi-tangent at $\alpha=\pi/2$ and $\alpha=3\pi/2$ for one of the quasi-shear waves. In other words an indentation is observed at these two points. This indentation produces an extra vertical tangent to the corresponding slowness curve in the case of a type II material (compare Fig. 4 with Fig. 5). The incidence angle corresponding to this tangent line is the extra critical angle that Ahmed and Rahman have noticed in their mathematical calculations. The value of this angle was verified for all materials examined in Ref. 2 from the corresponding slowness curves.

IV. CONCLUSIONS

In this paper, it was shown that the difference between type I and type II transversely isotropic materials is governed by the ratio c_{33}/c_{11}. If the normal stiffness of the cylinder in the isotropic plane is smaller/larger than the stiffness along the third axis, the material is of type I/II. A geometrical approach was also proposed for finding the extra critical angle, α_c, of a type II transversely isotropic material.

[1] F. Honarvar and A. N. Sinclair, "Acoustic wave scattering from transversely isotropic cylinders," J. Acoust. Soc. Am. **100**, 57–63 (1996).
[2] F. Ahmed and A. Rahman, "Acoustic scattering by transversely isotropic cylinders," Int. J. Eng. Sci. **38**, 325–335 (2000).
[3] M. Musgrave, *Crystal Acoustics, Introduction to the Study of Elastic Waves and Vibration in Crystals*, Holden-Day Series in Mathematical Physics (Holden-Day, San Francisco, 1970).
[4] B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Melbourne, Florida, 1990), Vol. 2.

Time-domain pressure response of arrays with periodic excitation (L)

Jeffrey A. Ketterling[a] and Frederic L. Lizzi
Riverside Research Institute, 156 William Street, New York, New York 10038

(Received 17 November 2002; revised 11 April 2003; accepted 16 April 2003)

A time-domain method utilizing the spatial impulse response is proposed to compute beam patterns for array transducers where each array element may have a unique continuously periodic (CP) excitation. The method is useful for cases when an efficient time-domain approach is desired for the steady-state and transient results. The pressures for each array element at a point **P** in the sound field are computed over the lowest period T and an appropriate time shift is applied. The resulting pressures are summed to give the total pressure at **P**. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1580834]

PACS numbers: 43.20.Rz, 43.40.Rj, 43.35.Wa [LLT]

I. INTRODUCTION

The spatial impulse response (SIR) method provides a fast, efficient time-domain method to calculate the sound field of a transducer.[1] The SIR at a point $\mathbf{P}(\mathbf{r})$ in space is described by the temporal function $h(\mathbf{r},t)$. The pressure at **P** is found by the convolution $-\rho\partial v(t)/\partial t * h(\mathbf{r},t)$, where $v(t)$ represents a uniform surface velocity excitation of the transducer surface **S**. The nonzero region of $h(\mathbf{r},t)$ falls between the times, $t_{\min}=r_{\min}/c$ and $t_{\max}=r_{\max}/c$, where c is the speed of sound. These times correspond to the propagation paths of the nearest ($r_{\min}$) and farthest ($r_{\max}$) distances between **P** and the transducer. Analytic expressions for $h(\mathbf{r},t)$ have been derived for simpler geometries such as circular pistons,[1] rectangular surfaces,[2] triangular apertures,[3] and spherical shells.[4,5] Alternate methods for computing $h(\mathbf{r},t)$ also exist for more complicated geometries.[6–9]

Once $h(\mathbf{r},t)$ has been calculated, the excitation may be an arbitrary $v(t)$. While cases solved in the literature typically deal with impulse excitations, there is no restriction on the duration of the excitation, and a continuously periodic (CP) drive of any frequency content may be approximated with a long-duration wavetrain. To compute the pressure for an array, each drive signal needs to be long enough to ensure that all array elements contribute to the steady-state response at **P**. However, this is not optimal from a computational standpoint, and it would be preferable to minimize the durations of $h(\mathbf{r},t)$ and $v(t)$ prior to convolution. When the signals are appropriately minimized, the resulting time-domain pressure is a single steady state with on–off transients. This is all the information that is needed to find the net steady-state or transient pressure at **P** from the M-array elements.

In this article, a time-domain method is proposed to calculate the sound field of a transducer with M elements, where each element may have a unique CP drive. The time durations of the signals $h_m(\mathbf{r},t)$ and $v_m(t)$ are minimized prior to convolution (or circular convolution), and the appropriate phase delays are applied to find the single cycle or transient solution. Summing the pressure wave for each element gives the final pressure waveform $p(\mathbf{r},t)$.[6,10–12] The method is ideal to incorporate novel CP excitations into existing time-domain SIR models that only address pulsed excitations.

We first summarize the SIR method and show how the pressure field for a CP excitation of the source would be analytically computed. We then demonstrate how to efficiently compute the steady-state solution or transient components of $p(\mathbf{r},t)$ for a CP-driven source for the more practical case of a discrete time numerical implementation. The method is demonstrated on a spherically curved annular array.

II. THEORY

A. General SIR formulation

We briefly outline the SIR method following Arditi *et al.*[5] The SIR method is derived starting with the Rayleigh diffraction integral.[13] Strictly speaking, the Rayleigh integral only applies to a baffled piston, but it also provides an excellent approximation for weakly focused transducers.[4,5,13] A description of the transducer geometry is shown in Fig. 1. The pressure at a point $\mathbf{P}(\mathbf{r})$ in space radiated by a surface **S** is described by $p(\mathbf{r},t)=-\rho\,\partial\phi(\mathbf{r},t)/\partial t$, where ρ is the medium density and $\phi(\mathbf{r},t)$ the velocity potential. The velocity potential may be written in terms of a convolution[1,2] $\phi(\mathbf{r},t)=v(t)*h(\mathbf{r},t)$, where $v(t)$ is the normal surface velocity of **S**. The pressure may now be expressed as

$$p(\mathbf{r},t)=-\rho\frac{\partial v(t)}{\partial t}*h(\mathbf{r},t), \tag{1}$$

where the SIR is defined as

$$h(\mathbf{r},t)=\frac{1}{2\pi}\int_S\frac{\delta(t-\mathbf{d}/c)}{\mathbf{d}}d\mathbf{S}, \tag{2}$$

where $\delta(t)$ is the Dirac delta function, c is the speed of sound, and **d** is a distance between **S** and **P**.

One way to visualize how $h(\mathbf{r},t)$ is computed is to imagine **S** as a collection of point sources. At some time t_i, all of the points that are a distance ct_i from **P** contribute to the value of $h(\mathbf{r},t_i)$. A value for $h(\mathbf{r},t)$ exists for all times

[a]Electronic mail: ketterling@rrinyc.org

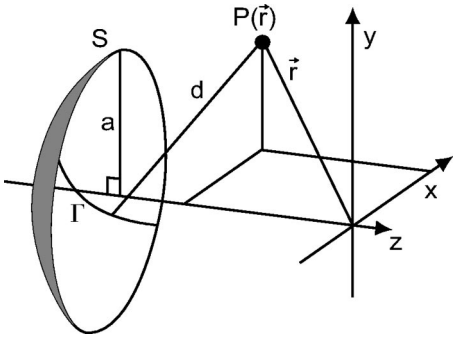


FIG. 1. Geometry used to calculate $h(\mathbf{r},t)$. The figure shows a spherical cap of radius a but the description may be used for any surface **S**. At some time t, a path Γ on the transducer surface is traced at a distance $\mathbf{d}=ct$ from **P**.

between $t_{\min}$ and $t_{\max}$ corresponding to the points on **S** that are the nearest and farthest from **P**. In most cases, the integration of Eq. (2) reduces to a single integral with an analytic or numerical solution. An exception to this rule would be at the focus of a spherically curved transducer where all the point sources arrive simultaneously and $h(\mathbf{r},t_i)$ is a function of the surface area of **S**.

B. Analytic description

For the case where $h(\mathbf{r},t)$ and $v(t)$ are known analytic functions, we want to find an effective method to compute $p(\mathbf{r},t)$ from Eq. (1). Each array element has its own drive signal $v(t)=g(\omega,t-t_1)$ with time delay t_1. The lowest frequency component of all the drive signals is $f_0=\omega_0/2\pi$ with period T. Because $p(\mathbf{r},t)$ will also have period T, the convolution of Eq. (1) only needs to be performed over one period. Any interval larger than T will provide duplicate information. The single-period solution for Eq. (1) becomes

$$p_{sp}(\mathbf{r},t)=-\rho\int_0^T \frac{\partial g(\omega,\tau-t_1)}{\partial t}h(\mathbf{r},t-\tau)d\tau, \tag{3}$$

where the solution is valid over the range $t=[0,T]$. Outside of this range the pressure is $p(\mathbf{r},t)=p_{sp}(\mathbf{r},jT)$ where j is the fractional portion of the ratio t/T. Because the drive signal is assumed to be CP for all t, there are no transient components to $p_{sp}(\mathbf{r},t)$.

To find the total pressure $p_{\text{tot}}(\mathbf{r},t)$ for an array of M elements where the pressure from ring m is $p_m(\mathbf{r},t)$, the single-period pressure contributions from Eq. (3) are summed to give

$$p_{\text{tot}}(\mathbf{r},t)=\sum_{m=1}^{M} p_m(\mathbf{r},t);t=[0,T]. \tag{4}$$

In most practical applications, the forms of $h(\mathbf{r},t)$ and $v(t)$ are discrete-time signals requiring a numerical approach.

C. Numerical implementation

For a numerical implementation, $h(\mathbf{r},t)$ is computed at a numerical sampling rate Δt from a set of analytic expressions valid over certain intervals of time or from a numerical integration of Eq. (2). A numerical sampling rate of $\geqslant 100$ times the highest frequency component of the drive signal is usually sufficient.[5] For a CP excitation, the numerical form of $v(t)$ must be expressed as a wavetrain of long but finite time duration. The finite duration $v(t)$ will therefore lead to transient components after convolution with $h(\mathbf{r},t)$. Ideally, we would like to minimize the duration of $v(t)$ and calculate just a single period of $p(\mathbf{r},t)$ while initially ignoring the transient components of the signal.

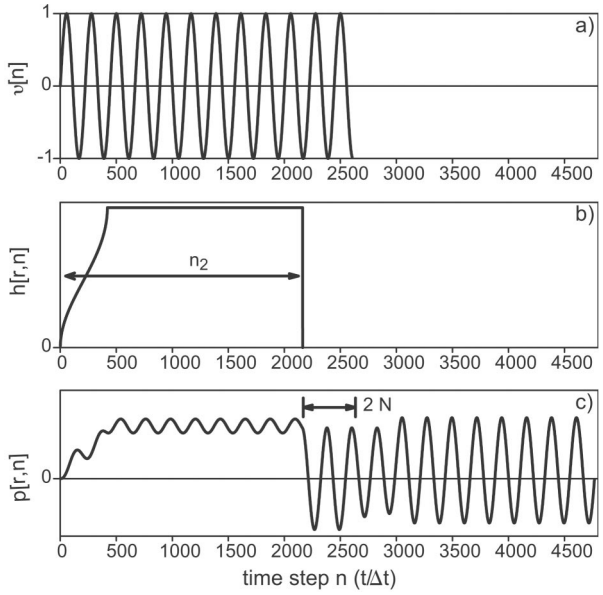


FIG. 2. Graphical example of the convolution of Eq. (1) for an arbitrary drive signal. h is a typical SIR of duration n_2 for a ring of an annular array (b). The duration of $v[n]$ is $2+n_2/N$ cycles (a). The resulting convolution has $2n_2+2N$ points, of which $2N$ represents a steady-state solution starting at point n_2 (c).

We first change to a discrete time notation where the pressure is denoted as $p[\mathbf{r},n]$ and $n=t/\Delta t$ represents an integer multiple of a time step Δt. The SIR is denoted as $h[\mathbf{r},n]$, where the nonzero duration of $h[\mathbf{r},n]$ is n_2 and $n_{\max}=n_{\min}+n_2$ ($n_{\max}=t_{\max}/\Delta t$ and $n_{\min}=t_{\min}/\Delta t$). For numerical calculations, it is more convenient to express the SIR in a time-shifted form $h[\mathbf{r},\hat{n}]$ where $\hat{n}=n-n_{\min}$. An example of a time-shifted SIR is shown in Fig. 2(b).

The drive signal is described as

$$v[\overline{n}]=g[\omega,\overline{n}\Delta t](u[\overline{n}]-u[\overline{n}-(n_2+2N)]), \tag{5}$$

where $u[n]$ is the unit step function and N represents one period $T/\Delta t$. The drive signal has a time shift of n_1 with $\overline{n}=n-n_1$. The duration of the drive signal represents $2+n_2/N$ periods. Figure 2(a) shows an example of $v[\overline{n}]$ for an arbitrary CP signal that is two periods longer than the duration of $h[\mathbf{r},\hat{n}]$ in Fig. 2(b). As will be demonstrated below, $2+n_2/N$ periods are the minimum number of periods needed to conveniently obtain a solution. Equation (1) is now expressed as a discrete time signal

$$p[\mathbf{r},\tilde{n}]\propto\sum_{l=0}^{2N+2n_2}\frac{\partial v[l]}{\partial t}h[\mathbf{r},\tilde{n}-l], \tag{6}$$

where $\tilde{n}=n-n_{\min}-n_1$. The convolution of Figs. 2(a) and (b) may be numerically carried out as a standard convolution or a circular convolution, with the result shown in Fig. 2(c).

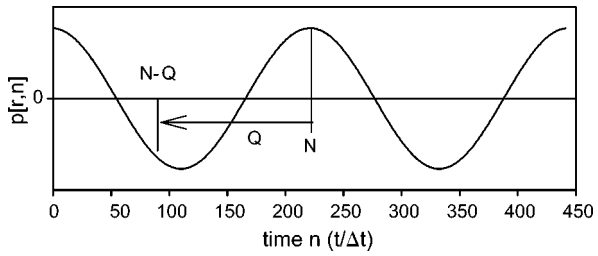


FIG. 3. The start of the two steady-state periods provides a new zero point. The single-period solution is referenced to the start time $n=0$ and spans from point $N-Q$ to $2N-Q$.

The first n_2 points of $p[\mathbf{r},\tilde{n}]$ represent the transient turn on interval of the signal. Then, there are two steady-state periods followed by n_2 additional transient points. The two steady-state periods

$$p[\mathbf{r},b]=p[\mathbf{r},\tilde{n}-n_2];\quad b=[0,2N], \tag{7}$$

are of interest (Fig. 3) for the CP signal.

We choose two periods rather than one because of a final time shift needed to account for the delays in $\hat{n}$, $\overline{n}$, and the transient duration n_2. This time shift will reference the phase back to $n=0$ at the transducer surface. This final step is essential for maintaining the appropriate phase delay between array elements. The time shift is $Q=\mathrm{Rem}[(n_1+n_2+n_{\min})/N]$, where Rem is the integer remainder of the ratio of the bracketed term. The total time shift may sum to several cycles, but only the relative phase shift over one cycle is relevant. The single-period solution is then

$$p[\mathbf{r},n]=p[\mathbf{r},b-(N-Q)];\quad n=[0,N], \tag{8}$$

and for an M element array the single-period solutions for all elements are summed to give the total pressure

$$p_{\mathrm{tot}}[\mathbf{r},n]=\sum_{m=1}^{M} p_m[\mathbf{r},n];\quad n=[0,N]. \tag{9}$$

If the transient components are of interest, then the results of Eq. (6) for the M-array elements may be used directly. Let $o_{\min}$ be the minimum value of $n^m_{\min}+n^m_1$. This represents the first of the m turn-on transients to arrive at **P**. Let $o_{\max}$ be the maximum value of $n^m_{\min}+n^m_1+n^m_2$. This represents the end of the final turn-on transient to arrive at **P**. The total turn-on transient signal at **P** is found by summing the signals $p_m[\mathbf{r},n-o_{\min}]$ on $n=[o_{\min},o_{\max}]$ where each signal p_m is assembled from $n^m_{\min}+n^m_1-o_{\min}$ zeros, then the n^m_2 points of the turn-on transient, and then the steady-state cycle N^m is repeated until $n=o_{\max}$. The turn-off transient is found in a similar fashion, but the full duration of the actual CP drive signal $v_m[n]$ must be considered before finding the appropriate $o_{\min}$ and $o_{\max}$. The signals are now assembled in a reverse sequence (steady-state periods, then the turn-off transient, and then zeros) prior to summing $p_m[\mathbf{r},n-o_{\min}]$.

III. SIMULATION OF AN ANNULAR ARRAY

To demonstrate the above numerical method, we apply it to an eight-ring spherically focused annular array. The equations to find $h(\mathbf{r},t)$ at any point **P** are found in Table I of Ref. 5. The radius of curvature R of the annular array is 9 cm and the diameter is 8 cm. The array has eight equal area annuli with no kerf between rings. The rings are located at radii of $a=1.45$, 2.04, 2.49, 2.87, 3.19, 3.49, 3.75, and 4 cm, where 0 cm is the center of the transducer and 4 cm is the outer edge. The frequency of excitation is 4.5 MHz and the time step used to numerically sample the signals is $\Delta t=10^{-9}$ s. A single frequency drive is used for simplicity and permits comparison to O'Neil's[13] classical result for the axial pressure of a focused transducer.

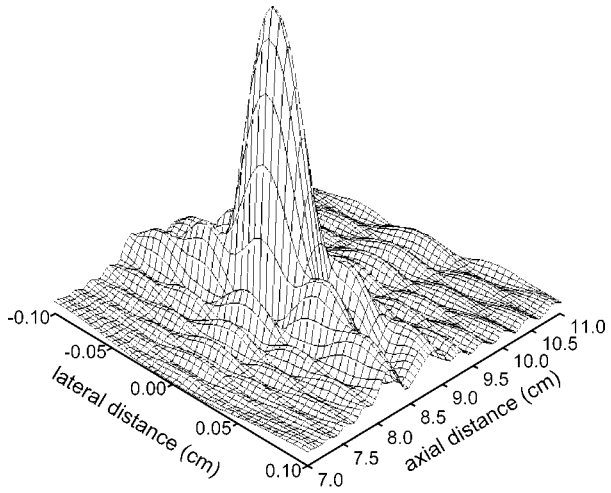


FIG. 4. Two-dimensional pressure-field distribution for an eight ring annular array focused at 8.5 cm. The geometric focus is at 9 cm. The vertical axis represents peak acoustic pressure.

The time delay of $v_m(t)$ on ring m required to shift the focus to an axial distance z is $t_1=a^2_m(1/R-1/z)/2c$, where $c=1500$ m/s.[5] The mean radius for each annulus is $a_m=\sqrt{(a^2_{\min}+a^2_{\max})/2}$, where $a_{\max}$ and $a_{\min}$ are the maximum and minimum radii of ring m. The geometric focus is at $z=9$ cm and corresponds to a delay of zero for all annuli. A delay of zero also makes the array equivalent to a single-element focused transducer. When the axial pressure is computed for the array with zero delay, the result is equivalent to that of O'Neil,[13] as was previously noted in Ref. 5.

Figure 4 shows an example of the beam profile with an electronic focus at $z=8.5$ cm. The 3-dB lateral beamwidth in this case is 370 μm, slightly less than the beamwidth of 390 μm at the geometric focus. For the results shown here, no apodization is used.

IV. CONCLUSION

A time-domain method to efficiently compute the pressure field from a CP-excited array using the SIR of each element was presented. The method is applicable to any numerical model capable of generating the time-domain SIR for an array element at a point **P** in space. For the steady-state solution, a single drive period with appropriate phase shift is computed for each array element m. The single-period pressures for all M elements are then summed to give the pressure at the point **P**. The turn-on and turn-off transients of the array may also be found directly from the convolution results of Eq. (6).

The modeling of CP-driven arrays is relevant for medical applications such as high-frequency focused ultrasound

(HIFU). HIFU applications typically utilize drive signals of finite duration (1–60 s) but high frequency (0.5–5 MHz). For modeling purposes, the sound field is continuous. Typically, an array is excited to generate a high acoustic intensity and induce hyperthermia and thermal necrosis in tissue at the focal region of the beam.[14] Unlike a single-element transducer, a large tissue volume may be treated with an array without the need to mechanically scan the transducer.[15–17] The methods proposed in this report will facilitate in the modeling of the electronic focus shift and the two-dimensional beam profiles of array transducers excited with novel CP drives.

ACKNOWLEDGMENTS

The authors wish to thank Cheri Deng and Dan Reuster for many useful discussions. Portions of this work were supported by NIH Grants EY01212 and CA84588.

[1] P. Stepanishen, ''Transient radiation from pistons in an infinite planar baffle,'' J. Acoust. Soc. Am. **49**, 1629–1683 (1971).
[2] J. Lockwood and J. G. Willette, ''High-speed method for computing the exact solution for the pressure variations in the nearfield of a baffled piston,'' J. Acoust. Soc. Am. **53**, 735–741 (1973).
[3] J. A. Jensen, ''Ultrasound fields from triangular apertures,'' J. Acoust. Soc. Am. **100**, 2049–2056 (1996).
[4] P. Penttinen and M. Luukkala, ''The impulse response and pressure nearfield of a curved ultrasonic radiator,'' J. Phys. D **9**, 1547–1557 (1976).
[5] M. Arditi, F. S. Foster, and J. W. Hunt, ''Transient fields of concave annular arrays,'' Ultrason. Imaging **3**, 37–61 (1981).
[6] J. A. Jensen and N. B. Svendsen, ''Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).
[7] B. Piwakowski and B. Delannoy, ''Method for computing spatial pulse response: Time-domain approach,'' J. Acoust. Soc. Am. **86**, 2422–2432 (1989).
[8] B. Piwakowski and K. Sbai, ''A new approach to calculate the field radiated from arbitrarily structured transducer arrays,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 422–440 (1999).
[9] J. F. Theumann, M. Arditi, J. J. Meister, and E. Jaques, ''Acoustic fields of concave cylindrical transducers,'' J. Acoust. Soc. Am. **88**, 1160–1169 (1990).
[10] P. Wu and T. Stepinski, ''Spatial impulse response method for predicting pulse-echo fields from a linear array with cylindrically concave surface,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1283–1297 (1999).
[11] D. H. Turnbull and F. S. Foster, ''Beam steering with pulsed two-dimensional transducer arrays,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**, 320–333 (1991).
[12] P. Crombie, P. A. J. Bascom, and R. S. C. Cobbold, ''Calculating the pulsed response of linear arraya: Accuracy versus computational efficiency,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 997–1009 (1997).
[13] H. T. O'Neil, ''Theory of focusing radiators,'' J. Acoust. Soc. Am. **21**, 516–526 (1949).
[14] G. ter Haar, ''Acoustic surgery,'' Phys. Today **54**, 29–34 (2001).
[15] F. Dupenloup, J. Y. Chapelon, D. J. Cathignol, and O. A. Sapozhnikov, ''Reduction of the grating lobes of annular arrays used in focused ultrasound surgery,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 991–998 (1996).
[16] C. A. Cain and S. I. Umemura, ''Concentric-ring and sector-vortex phased array applicators for ultrasound hyperthermia,'' IEEE Trans. Microwave Theory Tech. **34**, 542–551 (1986).
[17] T. Fjield, X. Fan, and K. Hynynen, ''A parametric study of the concentric-ring transducer design for MRI guided ultrasound surgery,'' J. Acoust. Soc. Am. **100**, 1220–1230 (1996).

Reciprocity and the active control of distributed acoustic sources (L)

Stephen J. Elliott[a] and Lewis Rees
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton S017 1BJ, United Kingdom

(Received 18 November 2002; accepted for publication 7 April 2003)

The velocity distributions of a pair of arbitrary acoustic sources are approximated by arrays of elementary sources, and a reciprocal relationship is established between the generalized force on one source and the total velocity of the other source. When the two distributed sources are individually constrained to vibrate with a single phase and are used for active sound control, it is shown that the attenuation in sound power from the pair of sources is the same regardless of whether one source velocity is fixed and the velocity of the other source is adjusted to minimize the power output, or vice versa. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1577559]

PACS numbers: 43.50.Ki [KAC]

I. INTRODUCTION

The acoustic effect of minimizing the total radiated sound power from a compact primary source using monopole secondary sources is well known.[1] In this paper we consider the minimization of sound power from a distributed primary source using a distributed secondary source. These distributed sources are assumed to be the limiting case of a large array of elemental sources, divided up into individual primary and secondary source arrays.

Within this framework a reciprocal relationship is established between the generalized force on the first source due to the velocity of the second and the generalized force on the second source due to the velocity of the first. This result is similar to the general reciprocal relation for coupling between the modes of two bodies described by Chertock.[2] An additional assumption is then made that although each of the two source distributions can vibrate with a nonuniform distribution of amplitudes, the vibration of each element in an individual source is entirely in-phase or 180° out of phase. An expression for the total sound power output of these two sources is then derived, which is quadratic in the real and imaginary parts of the velocities of both sources. By considering the velocity of the first source as fixed, i.e., it is acting as the primary source, the total power output is minimized by adjusting the real and imaginary parts of the velocity of the second source. Then the velocity of the second source is fixed, so that this now acts as the primary, and the total power output is minimized by adjusting the real and imaginary parts of the velocity of the first source.

The surprising result is that the attenuation in the sound power output is the same in both cases.

II. RECIPROCITY FOR DISTRIBUTED SOURCES

Assume that an array of pistonlike elements, each of which is small compared with the acoustic wavelength, is distributed within a medium in which sound transmission is reciprocal. The vector of complex harmonic forces acting on each element, $\mathbf{f}$, in response to the complex velocities of each element, vibrating at a single frequency and contained in the vector $\mathbf{u}$, is given by

$$\mathbf{f}=\mathbf{Z}\mathbf{u}, \tag{1}$$

where $\mathbf{Z}$ is the matrix of complex mechanical radiation impedances between each of the elements. The matrix is symmetrical since the sound transmission is assumed to be reciprocal, i.e., linear and with no net flow, as discussed, for example, by Fahy.[3]

We now assume that the total source distribution is divided up into two arrays, denoted by the subscripts A and B, so that

$$\mathbf{u}=\mathbf{s}_A u_A+\mathbf{s}_B u_b\,, \tag{2}$$

where u_A and u_B are complex scalar velocities associated with the two arrays and $\mathbf{s}_A$ and $\mathbf{s}_B$ will be called the shading vectors for the two arrays. Without loss of generality these vectors can be normalized such that the sum of their elements is unity, so that u_A and u_B correspond to the sum of the velocities of the elements in each array and will be termed the total velocities of the arrays. If all the elements in the arrays have equal areas, the volume velocity of an array is proportional its total velocity. Two total, or generalized, forces are now defined, which are the sum of the forces on each of the elements, weighted by the shading vectors above, i.e.,

$$f_A=\mathbf{s}_A^{\mathrm{T}}\mathbf{f} \tag{3}$$

and

$$f_B=\mathbf{s}_B^{\mathrm{T}}\mathbf{f}. \tag{4}$$

Using Eqs. (1) and (2), these forces can be written in the form

$$f_A=Z_{AA}u_A+Z_{AB}u_B\,, \tag{5}$$

$$f_B=Z_{BA}u_A+Z_{BB}u_B\,, \tag{6}$$

where $Z_{AA}=\mathbf{s}_A^{\mathrm{T}}\mathbf{Z}\mathbf{s}_A$, $Z_{AB}=\mathbf{s}_A^{\mathrm{T}}\mathbf{Z}\mathbf{s}_B$, $Z_{BA}=\mathbf{s}_B^{\mathrm{T}}\mathbf{Z}\mathbf{s}_A$, and $Z_{BB}=\mathbf{s}_B^{\mathrm{T}}\mathbf{Z}\mathbf{s}_B$. Since Z_{AB} is a scalar, it must be equal to its transpose, i.e., $\mathbf{s}_B^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{s}_A$, and since $\mathbf{Z}=\mathbf{Z}^{\mathrm{T}}$, because of the overall

[a]Electronic mail: sje@isvr.soton.ac.uk

 0001-4966/2003/114(1)/52/3/$19.00

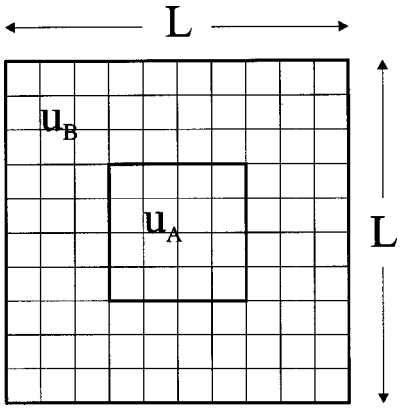


FIG. 1. Example of an array of 10×10 elemental sources divided into two distributed sources with velocities u_A and u_B.

reciprocity of the array, then the generalized reciprocal relationship can be established that

$$Z_{AB}=Z_{BA}. \tag{7}$$

To take a specific example, $\mathbf{s}_A$ and $\mathbf{s}_B$ may define the velocity distributions of the two separate bodies illustrated in Fig. 1. In Fig. 1 a 10×10 array of square piston elements is divided up into an inner source comprising the center 4×4 elements of the array, with total velocity u_A, and an outer source comprised of the three-element wide strips around the inner source, with total velocity u_B. In the simulations below, the array is assumed to be mounted in an infinite baffle and to radiate into an anechoic environment. If the number of elements in the array tends to infinity, the velocity distributions will become continuous. The point to point reciprocity theorem was extended by Smith[4] and Chertock[2] to derive a relation between the vibration of a body moving in a single mode and the radiated sound pressure, and the generalized force on that mode due to a monopole source. Chertock also mentions the extension of this result to the coupling between individual modes of two separate bodies. Equation (7) can be seen to be a statement of this result in the case where the modes of the two bodies are interpreted as generalized velocity distributions specified by the vectors $\mathbf{s}_A$ and $\mathbf{s}_B$. An advantage of the formulation above is that irregular source distributions, which could not be treated analytically, can be accurately approximated using an array of elemental sources and their coupling properties evaluated using simple matrix algebra.

III. MINIMUM POWER OUTPUT OF A PAIR OF DISTRIBUTED SOURCES

We now restrict the two source arrays to have arbitrary amplitude distributions but each to operate with a single phase, so that the shading vectors $\mathbf{s}_A$ and $\mathbf{s}_B$ are entirely real. Note that some elements of $\mathbf{s}_A$ and $\mathbf{s}_B$ could still be negative, indicating that these are operating exactly out of phase with the other elements, but no other phase difference is permitted.

The total sound power output of the overall array can be written as

$$\mathbf{W}=\tfrac{1}{2}\mathrm{Re}[\mathbf{u}^H\mathbf{f}], \tag{8}$$

where Re[ ] denotes the real part of the quantity in brackets and ^{H} denotes the Hermitian, complex conjugate, transpose. Using Eq. (2) and assuming that $\mathbf{s}_A$ and $\mathbf{s}_B$ are real, this can be written as

$$W=\tfrac{1}{2}\mathrm{Re}[u_A^*\mathbf{s}_A^{\mathrm{T}}\mathbf{f}+u_B^*\mathbf{s}_B^{\mathrm{T}}\mathbf{f}], \tag{9}$$

where $\mathbf{s}_A^{\mathrm{T}}\mathbf{f}$ and $\mathbf{s}_B^{\mathrm{T}}\mathbf{f}$ can be recognized as f_A and f_B, as in Eqs. (3) and (4), so that the power output becomes

$$W=\tfrac{1}{2}\mathrm{Re}[u_A^*f_A+u_B^*f_B]. \tag{10}$$

Using Eqs. (5) and (6), the power output may be written as

$$W=\tfrac{1}{2}\mathrm{Re}[u_A^*Z_{AA}u_A+u_A^*Z_{AB}u_B+u_B^*Z_{BA}u_A+u_B^*Z_{BB}u_B], \tag{11}$$

and since $Z_{BA}=Z_{AB}$, then

$$W=\tfrac{1}{2}[u_A^*R_{AA}u_A+u_A^*R_{AB}u_B+u_B^*R_{AB}u_A+u_B^*R_{BB}u_B], \tag{12}$$

where $R_{AA}=\mathrm{Re}[Z_{AA}]$, $R_{AB}=\mathrm{Re}[Z_{AB}]$, and $R_{BB}=\mathrm{Re}[Z_{BB}]$. The power output can now be expressed as the complex quadratic function[1]

$$W=x^*Ax+b^*x+x^*b+c, \tag{13}$$

where the complex independent variable x may be either u_A or u_B, with the constants A, b, and c chosen accordingly. Provided the real constant A is positive, then Eq. (13) has a unique global minimum when x is equal to

$$x_{\mathrm{opt}}=-A^{-1}b, \tag{14}$$

in which case the minimum value of W is

$$W_{\min}=c-b^*A^{-1}b. \tag{15}$$

The power output can thus be minimized by adjusting the velocity of one source, the secondary, if the velocity of the other source, the primary, is assumed to be held constant.

Let source A be the primary source in the first instance, in which case the velocity of the secondary source that minimizes the power output, $u_{B:\mathrm{opt}}$, divided by the velocity of the primary source, u_A, is given by

$$\frac{u_{B:\mathrm{opt}}}{u_A}=-\frac{R_{AB}}{R_{BB}}, \tag{16}$$

and the minimum power output of both sources, $W_{\mathrm{min}:B}$ divided by the power output of the primary source alone, $W_{\mathrm{primary}:A}$, is equal to

$$\frac{W_{\mathrm{min}:B}}{W_{\mathrm{primary}:A}}=1-\frac{R_{AB}^2}{R_{AA}R_{BB}}. \tag{17}$$

We now assume that source B is the primary source, in which case the velocity of the secondary source that minimizes the power output, $u_{A:\mathrm{opt}}$, divided by the velocity of the primary, u_B, is given by

$$\frac{u_{A:\mathrm{opt}}}{u_B}=-\frac{R_{AB}}{R_{AA}}, \tag{18}$$

and the minimum power output of both sources, $W_{\mathrm{min}:A}$, divided by the power output of the primary source alone, $W_{\mathrm{primary}:B}$, is

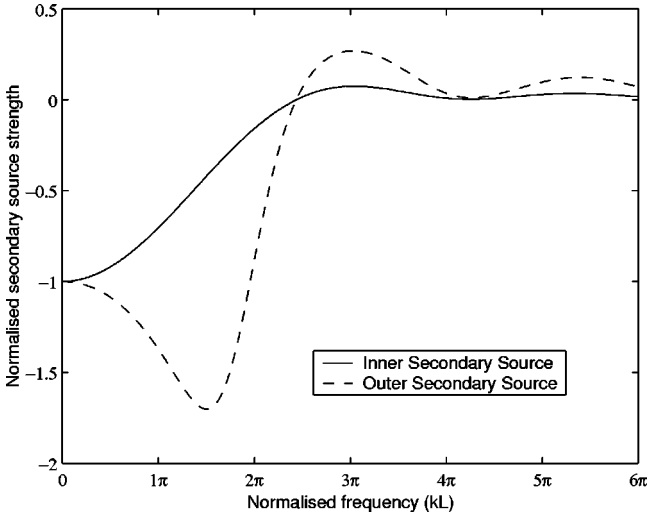


FIG. 2. Normalized velocity of the secondary source required to minimize the total power output of the array when the secondary source is either u_A (solid) or u_B (dashed), as a function of normalized frequency.

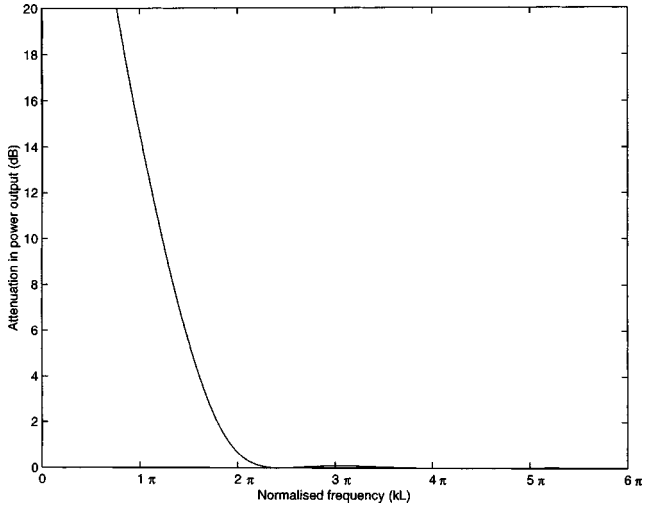


FIG. 3. Attenuation level in the power output of the array when either u_A or u_B acts as the secondary source, as a function of normalized frequency.

$$\frac{W_{\text{min:}A}}{W_{\text{primary:B}}} = 1 - \frac{R_{AB}^2}{R_{AA}R_{BB}}. \tag{19}$$

Notice that although the relative velocities of the secondary sources are different in the two cases, since in general R_{AA} is different from R_{BB}, the attenuations in the power output of the primary sources are the same regardless of whether one distributed source or the other is assumed to be the primary source.

This result is illustrated for the 10×10 element source array divided into the two distributions shown in Fig. 1. All the elements of $\mathbf{s}_A$ and all the elements of $\mathbf{s}_B$ are assumed to be equal, so that each of the elemental pistons in the two distributions vibrate with the same amplitude and phase. The normalized velocities of the secondary sources in the two cases, Eqs. (16) and (18), were calculated as a function of normalized frequency, kL, where k is the wave number and L is the dimensions of the outer array, and are plotted in Fig. 2. Notice that in both cases the total velocity of the secondary source is -1 times the total velocity of the primary source at low frequencies, so that the net volume velocity of the whole array is zero, as expected. Also, at high frequencies the velocity of the secondary source tends to zero in both cases, since little active control is possible. At other frequencies, however, the normalized velocity of the secondary source is very different in the two cases, with a much smoother transition from the low-frequency limit to the high-frequency limit for the case of the inner secondary source than for the case of the outer secondary source. The directivities of the two source distributions in the two cases are also significantly different in this frequency range. Figure 3 shows the level of attenuation in power output as a function of normalized frequency, which is identical for the two cases. The attenuation is negligible for $kL \geqslant 2\pi$, but considerable when $kL \ll \pi$. At low frequencies the primary source on its own radiates like a monopole, and the attenuation is found to be proportional to $(kL)^4$, so that the actively controlled source has a radiation efficiency which is the same as that of a quadruple source. These results do not change significantly over the frequency range plotted if a 20×20 array of elemental sources is divided up in a similar way to that shown in Fig. 1, which indicates that the array of elemental sources is behaving, to a close approximation, like a continuous distributed source.

IV. DISCUSSION AND CONCLUSIONS

The attenuation in sound power output for a given geometry of two source distributions in an active sound control system has been shown to be the same regardless of which is assumed to have a constant velocity and acts as the primary source, and which has its velocity adjusted to minimize the total power output, and thus acts as the secondary source. This is not obvious, since the ratio of the velocities of the secondary and primary sources are not the same in the two cases, nor are the directivities. The result can give some insight and reduce the computational burden when investigating suitable source distributions to best attenuate radiated sound power in active control systems.

It should be noted that although the source distributions are assumed to operate in phase, since the shading vectors are assumed to be real, they can have arbitrary shapes and even overlap. Complicated source distributions in three-dimensional space can be closely approximated by arrays of elemental sources using this formulation, which is also not restricted to free field conditions, provided the acoustic environment is linear and has no mean flow so that the normal reciprocal relationship between individual pairs of compact sources holds.

[1] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, New York, 1992).
[2] G. Chertock, "General reciprocity relation," J. Acoust. Soc. Am. **34**, 989 (1962).
[3] F. J. Fahy, "The vibro-acoustic reciprocity principle and applications to noise control," Acustica **81**, 544–558 (1995).
[4] P. W. Smith, "Response and radiation of structural modes excited by sound," J. Acoust. Soc. Am. **34**, 640–647 (1962).

Adaptive plasticity of loudness induced by chronic attenuation and enhancement of the acoustic background[a] (L)

C. Formby,[b] L. P. Sherlock, and S. L. Gold
Division of Otolaryngology—HNS and The University of Maryland Tinnitus and Hyperacusis Center, University of Maryland School of Medicine, Baltimore, Maryland 21201

(Received 20 December 2002; revised 25 April 2003; accepted 25 April 2003)

Strategies for treating hyperacusis, an anomolous condition of loudness perception, paradoxically seek either to minimize or enhance listeners' sound exposures. We show that these reciprocal treatment approaches, implemented with similar amounts of background sound attenuation and enhancement, produce diametrically opposite perceptual effects in normal listeners. Specifically, we show after continuous, 2-week earplugging and low-level noise treatments that listeners become more and less sensitive, respectively, to the loudness of sounds. This simple demonstration of adaptive plasticity is consistent with modification of a theoretical gain control process, which is the basis for desensitizing sound therapies used in treating hyperacusis and related sound tolerance problems. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582860]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Lj [MRL]

I. INTRODUCTION

A broad array of seemingly unrelated disorders and disease processes, some associated with hearing loss, share in common an anomoly of loudness perception known as hyperacusis.[1,2] This poorly understood problem is characterized by a remarkable intolerance for sound intensities that normally are judged to be comfortably loud.[3] Over the past decade, strategies for managing hyperacusis have changed diametrically from traditional sound-attenuating interventions that advocate sound avoidance[1,4] to desensitization protocols that promote exposure to an enriched sound background (i.e., sound therapy).[2,3] Proponents of desensitizing sound therapy champion the idea that a centrally mediated, compensatory gain control process is involved.[5,6] If so, then one would predict that normal loudness judgments also should be subject to modification by the reciprocal sound-attenuating and -enriching therapies variously used in the treatment of hyperacusis. We present evidence in this report that loudness judgments can be manipulated systematically, either upward or downward, in response to chronic increases or decreases in the normal listener's ambient sound background.

In practice, the use of sound therapy in the treatment of hyperacusis is achieved by prolonged bilateral exposure to soft noise, ostensibly, to desensitize a hypothetical gain control process.[5,6] This gain hypothesis is an intriguing, albeit unproven, idea that has been increasingly mentioned in both experimental[7–9] and clinical[2,5,10] research. Sound therapy assumes that, within the auditory pathways, there exists a compensatory gain process that regulates the supra-threshold sensitivity of the auditory system. Somehow this sensitivity has become set too high for the hyperacusis patient.[5] Furthermore, sound therapy assumes that the gain mechanism is plastic and its sensitivity can be reset to normal limits by continual exposure to enriched background sound.[2] Clinical evidence (i.e., post-treatment judgments of loudness discomfort levels shifted to higher sound intensities and accompanying subjective reports of improved sound tolerance) supports the benefits of sound therapy.[11–14] In contrast, it is alleged that prolonged use of sound-attenuating treatments creates greater problems of sound intolerance.[3,5] Although this idea seems defensible in terms of gain theory, inasmuch as diminished sound input potentially could sustain or even promote a higher gain setting, we know of no scientific evidence that either supports or refutes this claim.

In this study of normal loudness function, we evaluate a pair of unambiguous predictions that follow directly from theoretical assumptions about the role of the gain mechanism in control of hyperacusis and its treatment. The predictions are that an earplug (EP) treatment, which chronically attenuates sound input, will elicit a compensatory response that ''turns up'' the sensitivity of the gain process to enhance the loudness of sounds. In contrast, a noise instrument (NI) treatment that emits a soft noise, which chronically elevates ambient sound input, will evoke a compensatory response that ''turns down'' the sensitivity of the gain process to reduce perceived loudness. These treatment effects are described below and are shown to be consistent with both adaptive plasticity of loudness and the corresponding gain theory predictions.

II. METHODS

A. Participants

Ten consenting volunteers (mean age = 28.8 years; s.d. = 7.5) participated in this investigational study, which was approved by the IRB at the University of Maryland—Baltimore. All subjects were screened for normal hearing sensitivity (<20 dB HL across the audiometric frequency

[a]Portions of this work were presented in ''Plasticity of loudness perception,'' 143rd meeting of the Acoustical Society of America, Pittsburgh, PA, June 2002, and in ''Adaptive recalibration of chronic auditory gain: interim findings,'' Proceedings of the 7th International Tinnitus Seminar, Freemantle, Western Australia, March 2002.

[b]Electronic mail: cformby@smail.umaryland.edu

range from 250 to 8000 Hz) and normal loudness perception (pure-tone loudness discomfort level judgments in the range between 90 and 100 dB HL). Five of the ten participants were evaluated for both the NI and EP treatments, with at least 1 week separating the end of the first treatment and the onset of the second treatment. (On average, this time period was sufficient for loudness judgments, which were measured 1 week after termination of a given treatment, to recover to levels that were statistically equivalent to those for corresponding pretreatment baseline judgments.) Altogether, seven subjects were fitted with the NI treatment and eight subjects were fitted with the EP treatment. All subjects were able to continue their normal daily activities over the course of each treatment period.

B. Sound treatments

The EP treatment was achieved by fitting each subject bilaterally with a matched pair of earplugs (either Pacific Coast Labs, Sound Waves II plug with 3300 Ohm filter/adapter or Etymotic Research, custom modified E-A-R plug). The sound-enhancing NI treatment was accomplished with a matched set of open-canal, in-the-ear devices (General Hearing Instruments, model Tranquil). These devices produced a low-intensity noise over an effective range of frequencies between ~1000 and 8000 Hz, with a peak level of ~50 dB SPL near 6000 Hz. At the fitting appointment the outputs of the bilateral NI treatment were set to produce equal loudness between the ears, and the volume settings remained unchanged thereafter. Subjects wore each treatment continuously for at least 23 h each day over a 2-week treatment period. This period of time has been shown to produce measurable treatment effects in response to chronic unilateral sound attenuation.[15] Participants were permitted to remove their treatments once daily for cleaning, hygiene, and necessary maintenance.

The sound-attenuating and -enriching treatments were selected to produce similar patterns of audibility threshold shift. To assess these audibility treatment effects, pure-tone air-condition thresholds were measured for each subject at 500, 1000, 2000, and 4000 Hz at the onset and termination of each treatment. Measurements were performed according to standard clinical audiometric protocol under circumaural headphones with each treatment in place and with each treatment removed from the ears. (In addition, both sets of measurements were repeated at 1 week into each treatment to verify treatment stability.) From these data, the effective audibility threshold shifts were calculated for each treatment (see Fig. 1). The strength of both treatments was greatest above 1000 Hz, and, except at 4000 Hz, the ±1 standard deviation distributions of the mean audibility threshold shifts overlapped across frequency.

C. Loudness scaling protocol

Because the audibility threshold shifts produced by both sound treatments were greater at higher frequencies (see Fig. 1), we had the opportunity to evaluate frequency-specific treatment effects on the loudness judgments. We therefore measured loudness judgments separately with low- (500-Hz) and high- (2000-Hz) frequency tones. In the latter case, we selected 2000 Hz rather than 4000 Hz (the frequency associated with the largest audibility threshold shifts) for evaluation because the strengths of the treatments were better matched at 2000 Hz.

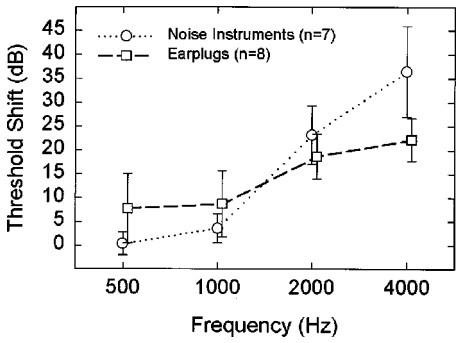


FIG. 1. Frequency-dependent strengths of the sound treatments inferred from average audibility threshold shifts (±1 standard deviation) produced by the noise instrument and earplug treatments.

Loudness judgments were measured using the Contour Test of Loudness Perception.[16] Participants assigned one of seven loudness response categories, progressively increasing from very soft to uncomfortably loud, to each of a series of four, constant-level, pulsed (200-ms), warble (5% frequency-modulated) tones. The test stimuli were produced by an audiometer (GSI, model 10) and presented monaurally by insert earphone (Etymotic, model ER-3A).

The Contour Test protocol was initiated by presenting the first series of warble tones at 20 dB HL. Each subject provided a numerical response to represent one of the loudness categories corresponding to that sound level (i.e., the response for this initial low sound level was routinely ''category 1,'' representing a judgment of ''very soft''). The subject's response was recorded by the examiner, who then increased the sound level by 5 dB (to 25 dB HL). The categorical response was recorded for this sound level, and so on. After each stimulus presentation and numerical response, the signal level was increased in 5-dB steps until a response was reported for ''category 7'' (i.e., representing a judgment of ''uncomfortably loud''). Subsequent to this response, the ascending sequence of presentation levels was repeated, beginning again from 20 dB HL. The loudness judgments for each sound level were repeated three or four times (depending upon response consistency) at each frequency. The estimated sound level for a given loudness response category was taken as the median value of levels assigned to that category.

The participants provided pretreatment baseline loudness judgments and loudness judgments at weekly intervals over the course of each 2-week treatment. The baseline judgments were measured immediately before each subject was fitted with one of the sound treatments. The treatment judgments were assessed in the same manner as at baseline, just after the treatments were removed from the ears and temporary threshold shifts were ruled out.

III. RESULTS

Group mean loudness judgments after 2 weeks of constant use of each treatment type are shown for each ear in

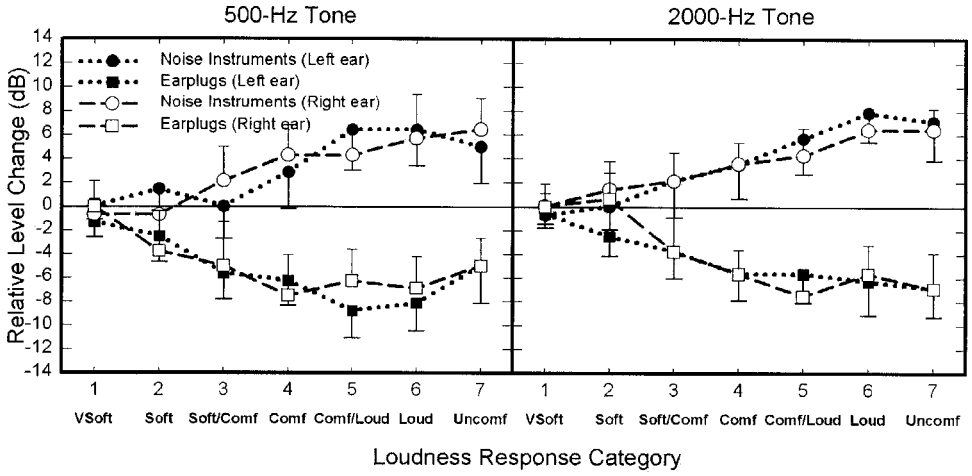


FIG. 2. Two-week treatment effects as a function of ear and stimulus frequency. Group mean changes (+ or −1 standard error bars) in categorical loudness judgments at termination of the noise instrument and earplug treatments for the left and right ears relative to pretreatment baseline sound levels (0 dB) measured for the 500-Hz (left panel) and 2000-Hz (right panel) warble tones. Baseline standard error of measurement (not shown) across all subjects, loudness categories, and test frequency conditions averaged 2.49 dB.

separate panels in Fig. 2 for the 500- and 2000-Hz tones relative to the corresponding pretreatment baseline loudness judgments. Typically, across all of the loudness response categories, and especially across categories judged to be comfortable and louder (i.e., categories 4–7), listeners needed more (+4 to +8 dB) and less (−5 to −9 dB) intense tones after the NI and EP treatments, respectively, to achieve the same loudness judgments as compared with baseline judgments. Thus, the resulting treatment effects in Fig. 2 are polar-opposite, mirror-image patterns of one another above and below the baseline judgments.

The loudness data for each treatment type are remarkably similar for both tones and, on average, the treatment effects are statistically the same for the two ears. We therefore collapsed the loudness data across ears to evaluate statistically the 2-week treatment changes from baseline for the NI and EP conditions as a function of loudness response category. These mean changes (in dB), which are shown with the corresponding *t*-test statistics and probability levels in Table I, were significantly different from baseline for categories 4–7 for the NI treatment and for categories 3–7 for the EP treatment at both test frequencies. Also shown in Table I are the overall mean differences in the treatment effects (collapsed across ears) as a function of loudness response category. These absolute treatment differences (in |dB|), which are shown with the associated Wilcoxon statistics and significance levels, were significantly different at both 500 and 2000 Hz for all but the two softest response categories (i.e., 1 and 2).

IV. DISCUSSION

The contrasting treatment effects reported here are consistent with predictions outlined in the Introduction. These

TABLE I. Two-week changes from baseline (in dB) for the noise instrument ($n=14$ ears) and earplug ($n=16$ ears) treatments and absolute treatment differences as a function of loudness response category for the 500- and 2000-Hz warble tones.

Loudness response category	Noise instrument minus baseline			Earplug minus baseline			Noise instrument minus earplug		
	Change (dB)	t^a	p	Change (dB)	t^a	p	Treatment difference |dB|	Z^b	p^c
					500 Hz				
1	−0.36	−0.43	0.3357	−0.63	−0.52	0.3046	0.27	0.23	0.4003
2	0.36	0.27	0.3965	−3.12	−1.62	0.0631	3.48	1.30	0.1025
3	1.07	0.56	0.2915	−5.31	−2.57	0.0106[d]	6.38	2.01	0.0267[d]
4	3.57	1.86	0.0429[d]	−6.88	−3.56	0.0014[d]	10.45	3.17	0.0018[d]
5	5.38	2.60	0.0110[d]	−7.50	−4.39	0.0003[d]	12.88	3.53	0.0007[d]
6	6.07	2.65	0.0101[d]	−7.50	−4.39	0.0003[d]	13.57	3.37	0.0011[d]
7	5.71	2.93	0.0059[d]	−5.00	−2.66	0.0090[d]	10.71	3.29	0.0013[d]
					2000 Hz				
1	−0.36	−0.56	0.2915	−0.31	−0.29	0.3869	0.05	0.267	0.3957
2	0.71	0.49	0.3175	−0.94	−0.68	0.2545	1.65	0.971	0.1698
3	2.14	1.15	0.1361	−3.75	−2.16	0.0238[d]	5.89	2.017	0.0265[d]
4	3.57	2.22	0.0225[d]	−5.63	−3.92	0.0007[d]	9.20	3.403	0.0010[d]
5	5.00	2.75	0.0082[d]	−6.56	−4.03	0.0006[d]	11.56	3.609	0.0006[d]
6	7.14	4.91	0.0002[d]	−5.94	−3.33	0.0023[d]	13.08	4.027	0.0002[d]
7	6.78	3.80	0.0011[d]	−6.88	−3.78	0.0009[d]	13.66	3.957	0.0002[d]

[a]Student's *t*-test statistic.
[b]Wilcoxon two-sample rank sum *Z*-test statistic.
[c]Probability level for one-sided *t*-test approximation of Wilcoxon *Z* statistic.
[d]Statistically significant treatment effect.

effects reveal direct evidence for adaptive plasticity of loudness, induced by changes in the listeners' chronic sound background, and provide indirect support for a compensatory gain process. Among the supporting evidence is the curious, but unambiguous, finding of frequency-invariant treatment effects. This invariance is remarkable in light of the relatively meager and stronger audibility threshold shifts (produced by both treatments) at 500 and 2000 Hz, respectively, and is especially surprising for the NI treatment, which produced virtually no threshold shift at 500 Hz (see Fig. 1), and yet affected loudness judgments similarly at both test frequencies. This peculiar property, together with the bilateral symmetry that characterizes these loudness data, seem consistent with a centrally mediated gain mechanism of the kind needed to explain bilaterally symmetric, frequency-invariant hyperacusis.[2]

A less provocative and perhaps less appealing explanation for this perceptual plasticity is that the listeners simply recalibrated their behavioral criteria for loudness in response to the chronic modifications in their listening backgrounds. For example, a listener who is accustomed to a quiet environment might reasonably set a more conservative criterion for loudness than one who is accustomed to louder conditions. If so, then we would expect corresponding changes in loudness judgments like those reported here for our EP and NI treatments. Circumstantial evidence consistent with this idea comes from large measurable declines in loudness discomfort level judgments of factory workers after their retirement from a noisy work setting.[17] This finding, however, also could be explained by gain recalibration in response to a reduced ambient noise floor after the workers were removed from the factory environment.

There are no data directly comparable to these results in the literature. Perhaps the most closely related evidence comes from loudness judgments measured immediately after removal of a chronically worn earplug. Florentine[15] reported that to obtain equal loudness matches between the ears it was necessary for her and her listeners to set the intensity lower in the ear from which the attenuating plug had just been removed relative to the sound intensity in the untreated ear. Such findings corroborate our bilateral EP treatment results, and indicate that loudness is enhanced after prolonged sound attenuation.

V. CONCLUSIONS

Whatever the ultimate explanation for these treatment effects, it seems clear that judgments of loudness are normally plastic and adaptable. These judgments somehow depend upon long-term changes in the ambient sound levels to which listeners are routinely exposed. Thus, this perceptual plasticity has important ramifications for everyday listening *vis a vis* adaptation to chronic variations within the ambient sound environment. One also can imagine manipulation and induction of psychological changes for other forms of perceived sensory magnitude (e.g., brightness, warmth, sweetness, etc.) by similar controlled modifications of the corresponding perceptual backgrounds. Most importantly, the treatment effects described here have direct implications for management of sound tolerance problems, which are common among hearing-impaired persons.[2] Such problems may prevent these individuals from successfully using hearing aids and cochlear implant devices to enhance communication. Efforts to characterize adaptive plasticity of loudness and to establish the mechanisms that control this intriguing perceptual plasticity in normal and impaired listeners await future studies.

ACKNOWLEDGMENTS

This investigation was supported by NIH research award Nos. R21DC04514 and K24DC00183 from the National Institute on Deafness and Other Communication Disorders. We gratefully acknowledge editorial assistance from Chinneta Pettaway-Willis, engineering support from Mead Killion and Andy Haapapuro in development of the EP treatment, statistical assistance from Charles Suter, and Monica Hawley's aid in revising the figures. We also thank Maureen Stone, Charles Watson, Moise Goldstein, David Robinson, Bert O'Malley, and Monica Hawley for constructive criticism of drafts of this manuscript.

[1] M. Anari, A. Axelson, A. Eliasson, and L. Magnusson, "Hypersensitivity to sound," Scand. Audiol. **28**, 219–230 (1999).
[2] C. Formby and S. L. Gold, "Modification of loudness discomfort level: evidence for adaptive chronic auditory gain and its clinical relevance," Sem. Hear. **23**, 21–34 (2000).
[3] J. Vernon, *Tinnitus: Treatment and Relief* (Allyn & Bacon, Boston, 1998).
[4] U. Katzenell and S. Segal, "Hyperacusis: review and clinical guidelines," Otol. Neurotol. **22**, 321–327 (2001).
[5] P. J. Jastreboff and J. W. P. Hazell, "A neurophysiological approach to tinnitus: clinical implications," Br. J. Audiol. **27**, 7–17 (1993).
[6] P. J. Jastreboff and J. W. P. Hazell, "Treatment of tinnitus based on a neurophysiological model," in *Tinnitus: Treatment and Relief*, edited by J. Vernon (Allyn & Bacon, Boston, 1998), pp. 201–217.
[7] G. M. Gerken, "Alteration of central auditory processing of brief stimuli: a review and a neural model," J. Acoust. Soc. Am. **93**, 2038–2049 (1993).
[8] C. Qui, R. J. Salvi, D. Ding, and R. Burkard, "Inner hair cell loss leads to enhanced response amplitudes in auditory cortex of unanesthitized chinchillas: evidence for increased system gain," Hear. Res. **139**, 153–171 (2000).
[9] R. J. Salvi, J. Wang, and D. Ding, "Auditory plasticity and hyperactivity following cochlear damage," Hear. Res. **147**, 261–274 (2002).
[10] P. J. Jastreboff and M. M. Jastreboff, "Tinnitus retraining therapy (TRT) as a method for treatment of tinnitus and hyperacusis patients," J. Am. Acad. Audiol **11**, 162–177 (2000).
[11] J. W. P. Hazell and J. B. Sheldrake, "Hyperacusis and tinnitus," in *Proceedings of the 4th International Tinnitus Seminar*, edited by J. M. Aran and R. Dauman (Kugler, Amsterdam, 1992), pp. 249–253.
[12] S. L. Gold, E. A. Frederick, and C. Formby, "Shifts in dynamic range for hyperacusis patients receiving tinnitus retraining therapy (TRT)," in *Proceedings of the 6th International Tinnitus Seminar*, edited by J. W. P. Hazell (Tinnitus & Hyperacusis Centre, London, 1999), pp. 297–301.
[13] S. L. Gold, C. Formby, E. A. Frederick, and C. Suter, "Shifts in loudness discomfort level in tinnitus patients with and without hyperacusis," in *Proceedings of the 7th International Tinnitus Seminar*, edited by R. Patuzzi (Univ. of Western Australia, Crawley, W.A., 2002), pp. 170–172.
[14] J. W. P. Hazell, J. B. Sherldrake, and R. L. Graham, "Decreased sound tolerance: predisposing factors, triggers and outcomes after TRT," in *Proceedings of the 7th International Tinnitus Seminar*, edited by R. Patuzzi (Univ. of Western Australia, Crawley, W.A., 2002), pp. 255–261.
[15] M. Florentine, "Relation between lateralization and loudness in asymmetrical hearing losses," J. Am. Aud Soc. **1**, 243–251 (1976).
[16] R. M. Cox, G. C. Alexander, I. M. Taylor, and C. A. Gray, "The contour test of loudness perception," Ear Hear. **18**, 338–400 (1997).
[17] W. Niemeyer, "Relations between the discomfort level and the reflex threshold of the middle ear muscles," Audiology **10**, 172–176 (1971).

Second derivative analysis of consonant-vowel transition waveforms (L)

K. H. Norian[a)]
Electrical and Computer Engineering Department, Lehigh University, 19 Memorial Drive West, Bethlehem, Pennsylvania 18015

(Received 26 February 2002; accepted for publication 26 February 2003)

The subject of this letter is the characterization of consonant-vowel syllables, through waveforms of the second derivative of the transition, to identify novel features within the transition. The second derivative of the first cycle of the transition region results in a modulated sinusoidal waveform with a spectral peak that is a function of the site of articulation of the consonant in the consonant-vowel syllable. The velars and alveolars give the lowest and highest spectral peaks, respectively, while the bilabial peaks occupy the mid-frequency range. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1569939]

PACS numbers: 43.72.Ar [DOS]

I. INTRODUCTION

The subject of the present work is to identify new acoustic cues within the transition region by using second-order differentiation. Whereas many modern speech analysis methods[1] and fundamental frequency extraction algorithms[2,3] analyze speech waveforms, in the present work sound waveforms are first transformed into their second derivative and the resulting waveforms are then analyzed. This method of characterizing speech waveforms has resulted in the identification of new, information-bearing features within transition regions. Whereas these features were not obvious in speech waveforms they could be seen in their second derivative transforms.

II. RESULTS AND DISCUSSION

Consonant-vowel syllables, consisting of the stop consonants /k,g,b,p,t,d/, each in combination with the vowel /a/, uttered by ten adult speakers (five female and five male) were digitized at 22-kHz sampling rate with 8-bit amplitude quantization. Twenty samples of each syllable were taken from each speaker. A virtual instrument was implemented, using graphical programming,[4] to process and display the results. Figure 1 shows a set of results that would appear on the display panel of the virtual instrument, where V is the output voltage of the microphone used to record the air pressure variations of the acoustic signal, and t is time. The waveform in Fig. 1(a) represents the sound intensity variation in the transition region for [ga] that starts with the first cycle, C1, and shows its development in time. The six waveforms in Figs. 1(b)–(g) show the second derivative of the transition regions for [ka], [ga], [ba], [pa], [ta], and [da], respectively, with each transition starting with the first cycle and including about three subsequent cycles. The dashed line in each of these six waveforms indicates the portion of the second derivative that corresponds to the first cycle in its respective transition region. These waveforms show that double differentiation transforms the first cycle of the sound waveform into a modulated sinusoidal. The spectrum of the second derivative of the first cycle of each CV transition region appears to the right of its respective waveform, for each of Figs. 1(b)–(g), with the spectral peak in each spectrum giving the fundamental frequency for its respective modulated sinusoidal.

The first cycle of the transition region was determined visually. The manner of determination was to start the inspection in the middle of the vowel, and to move left, to earlier time, until the end of the transition region was detected, i.e., at the time when the amplitude of cycle envelopes began to decrease (moving leftward). Then, by moving to still earlier time, towards the beginning of the transition region, the cycle with the smallest envelope amplitude was chosen as the first cycle. There were cases where the first cycle could not be clearly distinguished because it was difficult to precisely determine the start of the transition region, where signal energy was low, although later cycles of the transition region could be distinguished as signal energy increased (moving further to the right) and then reached a steady state. In these cases an approximate time range for the beginning of the transition region was estimated, and a sliding window was then used within this time range in the second differential waveform, to calculate the spectrum of adjacent regions within this range, until the region with the spectrum giving the single peak was found. The first such region in the range was declared the first cycle.

The spectral peaks for the subjects studied appear in Table I. Each speaker had her/his own unique set of spectral peaks with each peak associated with a different consonant in the CV transition. For each speaker spectral peaks were paired together; the syllables containing the velars /k,g/ had the lowest spectral peaks, those containing the alveolars /t,d/ had the highest, while those involving the bilabials /p,b/ had mid-frequency peaks. This grouping of the peaks indicates that the value of the spectral peak is a function of the site of articulation of the consonant.

These peaks were also present at the start of the transition region of CV transitions when the consonant was uttered in combination with /e/, /i/, /o/ or /u/. Table II shows the

a)Electronic mail: knh0@lehigh.edu

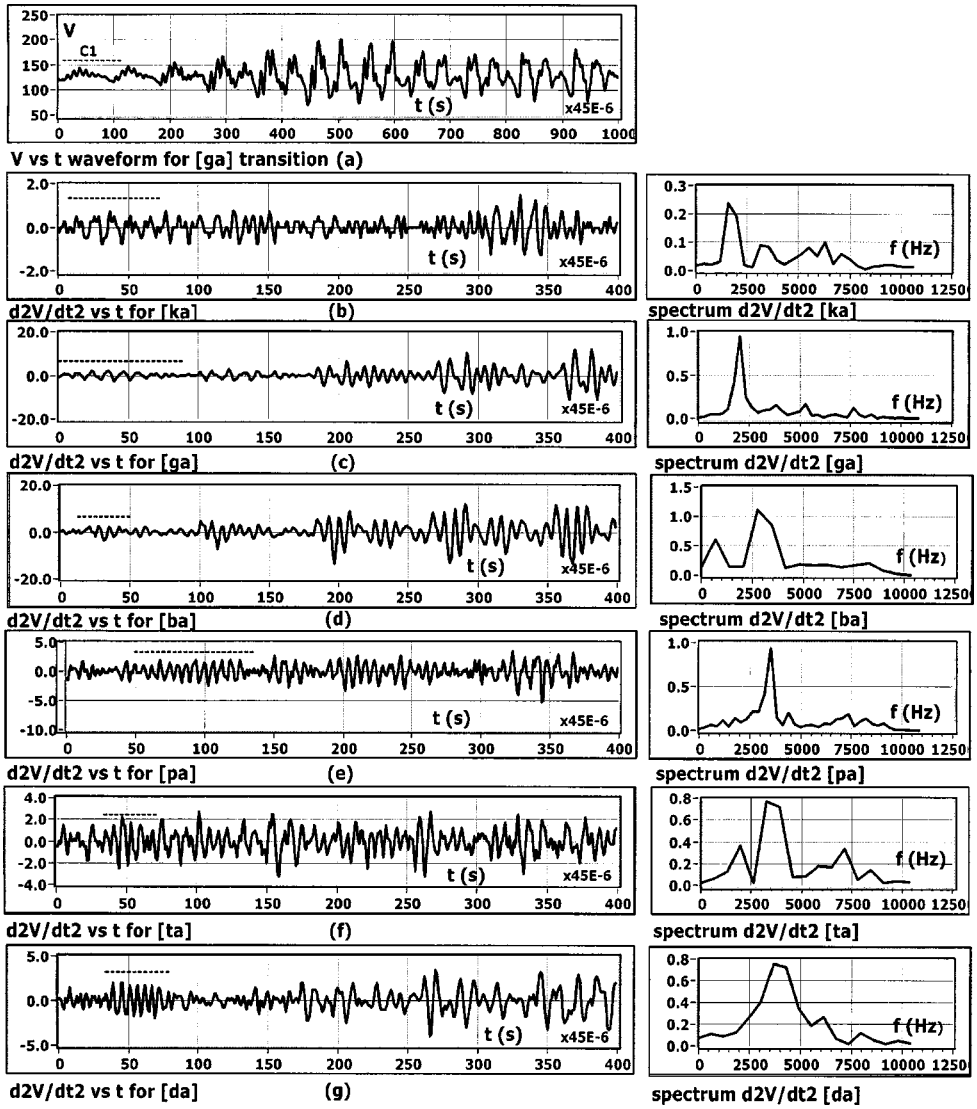


FIG. 1. (a) The variation of the syllable [ga] against time *t* starting with the first cycle, C1, of the transition region. (b)–(g) The variations of the second derivative for syllables [ka], [ga], [ba], [pa], [ta], and [da], respectively. To the right of each waveform in (b)–(g) is the frequency spectrum of the first cycle of the waveform. The dashed line in each waveform of the second derivative gives the duration of the first cycle.

spectral peaks for subject M1, of Table I, for the stop consonants in combination with each of the vowels. The results show that the peaks are the same for each consonant and are independent of the vowel in the CV transition. The spectral peaks obtained for the other nine subjects, with the stop consonants in combination with the vowels /e/, /i/, /o/ or /u/, were again the same as those given in Table I for the respective subject.

TABLE I. The spectral peaks f0, and the corresponding standard deviations (s.d.), of the second derivative of the first cycle of the transition region for each of the stop consonants in the context of /a/, for ten adult subjects. F=female and M=male.

	[ka]		[ga]		[pa]		[ba]		[ta]		[da]	
	f0	s.d.	f0	s.d.	f0	s.d.	f0	s.d.	f0	s.d.	f0	s.d.
F1	1640	90	1850	60	2480	50	2740	90	3800	60	4310	110
F2	1640	50	1870	100	2230	20	2590	60	3150	30	3980	80
F3	1520	30	1790	70	2400	60	2600	70	3870	80	4450	160
F4	1560	30	1750	90	2350	40	2670	50	3620	60	3810	90
F5	1690	50	1940	50	2440	60	2570	20	3550	90	4160	100
M1	1650	90	2030	60	2690	50	3100	80	3830	50	4180	140
M2	1460	30	2040	60	2660	90	2950	40	3450	90	3960	60
M3	1680	70	2010	100	2580	80	2720	70	3950	90	4300	160
M4	1480	30	2000	80	2730	60	3290	20	3410	70	3810	130
M5	1580	60	2030	110	2610	50	3300	30	3500	20	4050	70

TABLE II. The spectral peaks in Hz, with corresponding standard deviations, for the stop consonants in combination with each of the vowels for subject M1 of Table I.

[ka]	1650	90	[ke]	1690	50	[ki]	1660	60	[ko]	1630	50	[ku]	1580	50
[ga]	2030	60	[ge]	2090	60	[gi]	2070	50	[go]	2050	40	[gu]	2100	20
[pa]	2690	50	[pe]	2640	50	[pi]	2680	30	[po]	2570	50	[pu]	2700	30
[ba]	3100	80	[be]	3090	80	[bi]	3110	70	[bo]	3120	30	[bu]	3020	60
[ta]	3830	50	[te]	3850	110	[ti]	3890	90	[to]	3880	60	[tu]	3880	10
[da]	4180	140	[de]	4140	100	[di]	4200	50	[do]	4010	10	[du]	4080	70

The human nervous system uses double differentiation in lateral inhibition to improve signal resolution.[5–8] If it is assumed that double differentiation is part of the signal processing that occurs in the auditory cortex, then the position of the spectral peak may be an acoustic cue to identifying consonants. Moreover, enhancing the modulated sinusoidal for each syllable and then inverse transforming the waveform using double integration may be a way of speech enhancement.

Cycles subsequent to the single tone oscillation of the transform of the first cycle of the transition region are multitone waveforms where the fundamental frequency of the first cycle is also present. In the syllable [ga], for instance, the spectral peak of the first cycle increases in amplitude in up to five cycles beyond the first one. The equation of the waveform of the second differential of the first cycle and the contribution of this harmonic to subsequent cycles can then be given by

$$\frac{d^2V}{dt^2} = \frac{(t^m + t^{2m})}{10a^6} \sin(2\pi f t), \tag{1}$$

where f is frequency and a and m are constants. Figure 2 shows a plot of Eq. (1) for a normalized frequency of 0.46, and with $a = 0.39$ and $m = 0.82$. This equation is one of several that can be used to generate the oscillation seen in the first cycle and the development of this harmonic alone in subsequent cycles of the transition region. It is not a derived expression; it is the equation of a modulated sinusoidal that has been found to be useful in the LabVIEW signal processing environment to generate a family of curves, by varying a, f, and m, that can then be used in correlation studies to identify the oscillations, as one example of the many applications of the findings reported here.

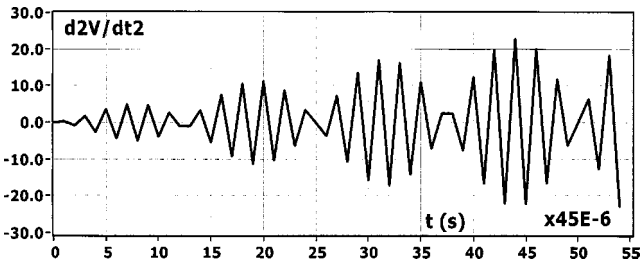


FIG. 2. Plot of Eq. (1) with $a = 0.39$, $m = 0.82$, and using a normalized frequency of 0.46. It shows the variation of the second derivative of the first cycle in the transition region of the syllable [ga], and the contribution of this harmonic to subsequent cycles of the transition region.

III. CONCLUSIONS

The second-order differential of the first cycle of the consonant-vowel transition results in a modulated sinusoidal with a spectral peak, which is a function of the site of articulation of the consonant in the consonant-vowel syllable. Future work aimed at enhancing these spectral peaks, through new signal processing algorithms, may be a way to improve speech intelligibility. This letter shows the importance of the method of double differentiation of signals as a signal processing tool for feature extraction in waveform analysis.

[1] R. D. Kent and C. Read, *The Acoustic Analysis of Speech*, 2nd ed. (Thomson Learning, Delmar, 2002).
[2] W. J. Hess, "Algorithms and devices for pitch determination of speech signals," Phonetica **39**, 219–240 (1982).
[3] V. Parsa and D. G. Jamieson, "A comparison of high precision F0 extraction algorithms for sustained vowels," J. Speech Lang. Hear. Res. **42**, 112–126 (1999).
[4] J. Y. Beyon, *LabVIEW Programming, Data Acquisition and Analysis* (Prentice-Hall, Upper Saddle River, NJ, 2001).
[5] F. Ratliff, H. K. Hartline, and W. H. Miller, "Spatial and temporal aspects of retinal inhibitory interactions," J. Opt. Soc. Am. **53**, 110–120 (1963).
[6] G. von Bekesy, *Sensory Inhibition* (Princeton U. P., Princeton, NJ, 1967).
[7] S. Deutsch, "Two-stage lateral inhibition for auditory selectivity," Bull. Math. Biol. **39**, 259–266 (1977).
[8] S. Deutsch and A. Deutsch, *Understanding the Nervous System* (IEEE, New York, 1993).

The effect of trabecular material properties on the frequency dependence of backscatter from cancellous bone (L)

Keith A. Wear[a)]
U.S. Food and Drug Administration, Center for Devices and Radiological Health, HFZ-142, 12720 Twinbrook Parkway, Rockville, Maryland 20852

(Received 22 July 2002; accepted for publication 27 December 2002)

Previous experimental studies indicate that backscatter coefficient for human calcaneal trabecular bone varies approximately as frequency cubed. This frequency dependence has been shown to be consistent with a model in which trabeculae are thought of as long thin cylinders composed of a substance with the same material properties as hydroxyapatite. The true material properties of human trabecular bone are not known however. Based on reported measurements of material properties of many bones and bonelike substances, it is possible that the density and longitudinal sound speed of trabecular bone material are far lower than the hydroxyapatite model would suggest. In this letter, it is shown that the frequency dependence of backscatter is still expected to be approximately cubic for wide ranges for density and longitudinal sound speed (encompassing the conceivable ranges for trabecular bone). [DOI: 10.1121/1.1554692]

PACS numbers: 43.80.Cs [FD]

I. INTRODUCTION

The investigation of scattering from trabecular bone is important for two reasons. First, it can help elucidate mechanisms responsible for attenuation (the combined result of absorption and scattering), which has been shown to have great diagnostic utility (Langton *et al.*, 1984; Rossman *et al.*, 1989; Zagzebski *et al.*, 1991; Glüer *et al.*, 1996; Hans *et al.*, 1996; Langton *et al.*, 1996; Bauer *et al.*, 1997; Bouxsein *et al.*, 1997; Chappard *et al.*, 1997; Laugier *et al.*, 1997a). Second, scattering measurements have shown some diagnostic promise in their own right in studies *in vitro* (Roberjot *et al.*, 1996; Wear and Armstrong, 2000; Hoffmeister *et al.*, 2000) and *in vivo* (Wear and Garra, 1997; Giat *et al.*, 1997; Laugier *et al.*, 1997b; Wear and Garra, 1998; Wear and Armstrong, 2001; Roux *et al.*, 2001). Backscatter provides information regarding size, shape, number density, and elastic properties of scatterers (Faran, 1951). In cancellous bone, trabeculae are likely candidates for scattering sites due to the high contrast in acoustic properties between mineralized trabeculae and marrow (Wear, 1999; Luo *et al.*, 1999). The diminished number and thicknesses of trabeculae that accompany aging and increased fracture risk would be expected to reduce backscatter.

In a previous investigation (Wear, 1999), measurements of average frequency-dependent backscatter from human calcaneal trabecular samples conformed fairly well to a power law dependence on frequency (throughout a typical diagnostic range: 300–700 kHz) with an exponent slightly greater than 3. (Due to the high attenuation coefficient of bone, the useful frequency band is much lower than that normally used for soft tissues.) A subsequent study reported a similar frequency dependence (Chaffai *et al.*, 2000). These findings have implications regarding the relative roles of scattering and absorption in determining attenuation in trabecular bone. Attenuation for trabecular bone in the 300–700-kHz range has been found in numerous studies to be approximately proportional to frequency to the first power. These two different coexisting frequency dependences could be consistent only if absorption is a much larger component of attenuation than scattering.

Empirical estimates of the exponent of the power law of frequency-dependent backscatter have been demonstrated to be consistent with a model in which trabeculae correspond to cylinders that are long (compared to the beam width), thin (relative to the wavelength), and oriented approximately perpendicular to the ultrasound propagation direction (Wear, 1999). Other models have also been shown to be useful for explaining experimental observations. One model assumes an isotropic statistically homogeneous scattering model, in which weak scattering structures are randomly distributed (Chaffai *et al.*, 2000). The other model portrays bone as a random continuum containing identical scatterers and assumes that scattering is proportional to the mean fluctuation in sound speed (Nicholson *et al.*, 2000).

Previously reported theoretical predictions (Wear, 1999) of frequency-dependent backscatter assumed that the properties of the trabecular bone material are the same as those for hydroxapatite, as measured by Grenoble *et al.* (1972). In reality, the true parameter values appropriate for trabecular bone are not known with great accuracy. There exist wide ranges of reported material property values for bone and bonelike substances. (See Table I.) Density and sound speed for hydroxyapatite are at the high ends of these ranges. This fact opens the previous study to the criticism that its conclusions were based on extreme parameter values which may differ considerably from the true values. The purpose of this letter is to explore the extent to which theoretically predicted

[a)]Electronic mail: kaw@cdrh.fda.gov

TABLE I. Material properties of bone and bonelike substances. The parameters listed are longitudinal sound velocity (c_l), shear velocity (c_s), density (ρ), and Poisson's ratio (ν).

Reference	Substance	c_l (m/s)	c_s (m/s)	ρ (g/cc)	ν
Hosokawa and Otani (1997)	Bovine cancellous bone	3800	2000	1.95	0.32
Anderson *et al.* (1998)	Hydroxyapatite	6790		3.22	0.28
Luo *et al.* (1999)	Cortical bone	2900	1303	1.85	
Gong *et al.* (1964)	Bovine cancellous bone			1.93	
Lang *et al.* (1970)	Bovine cortical bone			1.96	0.32
Williams (1992)	Bovine cancellous bone	3800[a]			
Rho *et al.* (1993)	Bovine cancellous bone	2898±85			
Ashman and Rho (1988)	Human cancellous bone	2639–2754		1.73–1.80	
	Bovine cancellous bone	2501		1.74	
Grenoble *et al.* (1972)[b]	Powdered human femur	3917	2020		0.33
	Fresh bovine femur	4890	2495		0.34
	Hydroxyapatite (mineral)	5565	4765		0.27
	Hydroxyapatite (synth.)	5628	4818		0.28

[a]Extrapolated (to zero porosity) from bulk measurements over a range of porosities (where porosity is the volume fraction of marrow). See Williams (1992, Fig. 2).
[b]Assuming ρ=1.96 gm/cc.

approximate cubic frequency dependence of backscatter may be extended to other assumed values for density and sound speed that may be more appropriate for trabecular bone.

II. METHODS

Faran developed a model for elastic scattering from a cylinder in a fluid (Faran, 1951). This model may be used to predict the intensity of backscatter from a cylindrical scatterer exposed to an incident plane wave. The axis of the cylinder is assumed to be oriented perpendicular to the acoustic propagation direction. Experimental measurements of scattering from a wire target suggest that Faran's expression is approximately proportional to the measured differential scattering cross section (from which the backscatter coefficient is derived) (Wear, 1999). Faran's model requires assumed values for the densities and longitudinal velocities for the scattering and embedding media. In addition, Poisson's ratio for the scattering medium must be specified. Finally, the diameter of the cylinder, the frequency of sound, and the angle between incident and observation directions are required.

In the present analysis, Faran's model was used to predict frequency-dependent backscatter. A wide range of assumed material properties for bone was employed. See Table I. Poisson's ratio seems to be fairly consistent for a variety of bone and bonelike substances and was taken to be 0.3. Trabecular diameter was assumed to be 120 μm (Hausler *et al.*, 1999). The embedding medium was assumed to be water (as is commonly used in *in vitro* experiments), with a longitudinal velocity of 1480 m/s and a density of 1 g/cc. For more details regarding computations, see Wear (1999).

Backscattered intensity (e.g., the response from a unit incident plane wave) $I(f)$ was assumed to obey a power law dependence on frequency, f, so that $I(f)=Af^n$ over a range of frequencies corresponding to the usable bandwidth of the transducer (300–700 kHz). Linear fits of the form $\log[I(f)]=\log A+n\log f$ were performed. The frequency dependence of backscatter was characterized by the exponent n.

III. RESULTS

Frequency-dependent backscatter, numerically obtained using the theory of Faran, for four combinations of assumed density and longitudinal sound speed (near the extremes from Table I) are shown in Fig. 1. These parameter values encompass the conceivable range for trabecular bone. It may be seen that over the frequency band of interest for experimental measurements (300–700 kHz), the frequency dependence of backscatter was similar for all parameter choices, despite the wide range of densities and longitudinal sound speeds employed.

In Fig. 2, exponents of power law fits to backscattered intensity versus frequency (over the range from 300 to 700 kHz) are shown as functions of the assumed value for longitudinal velocity of trabecular material. Over the plausible ranges for densities (1.7 g/cc$<\rho<$3.2 g/cc) and longitudinal velocities (2500 m/s$<c<$7000 m/s) for trabecular bone ma-

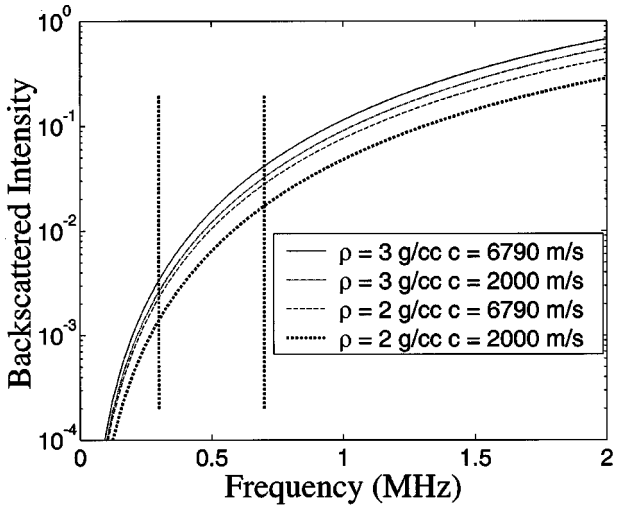


FIG. 1. Backscattered intensity from a cylindrical scatterer (diameter=120 μm) immersed in water for various choices of density and longitudinal sound velocity. The frequency dependence of backscatter is approximately the same for all pairs of parameter choices. The vertical dashed lines correspond to the frequency band of analysis for experimental measurements (300–700 kHz).

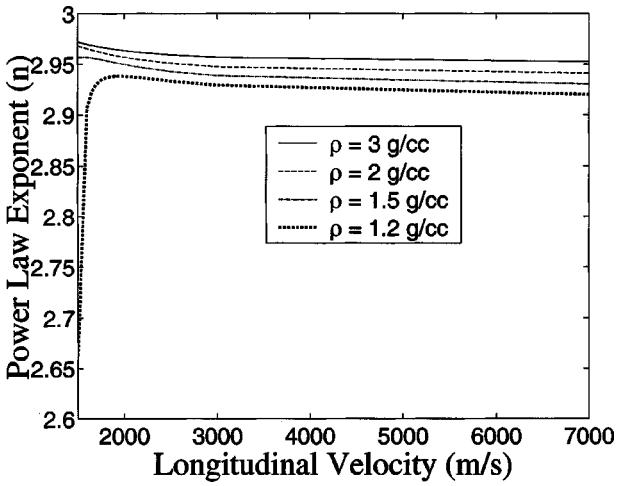


FIG. 2. The exponent of a power law fit (n) to backscattered intensity (I) versus frequency (f) from a cylindrical scatterer immersed in water. It has been assumed that $I(f)=Af^n$. The exponent of the power law fit is equal to the slope of the linear fit of log-transformed data ($\log[I(f)]=\log A+n\log f$). Fits were performed over the range from 300 to 700 kHz.

terial (see Table I), the exponent remained very close to 3 (ranging from 2.93 to 2.97). The exponent only began to deviate substantially from 3 when material values departed substantially from the expected range for bone and approached levels for water ($\rho<1.5$ g/cc and $c<1800$ m/s).

IV. CONCLUSION

Over a wide range of assumed values for density and sound speed for trabecular bone material, Faran's cylinder model predicts that backscatter varies approximately as frequency cubed, consistent with experimental measurements. So the conclusions of the previous study (Wear, 1999) are not restricted to a narrow choice of parameter values.

Measurements of power law exponent (n) for human calcaneus in the low-frequency range have been reported as 3.26 (Wear, 1999) and 3.38 (Chaffai *et al.*, 2000), which are a little higher than would be predicted by the cylinder model. The difference of 0.3–0.4 between theory and experiment may be due in part to the fact that trabecular bone has a more complex geometry than the cylinder model would suggest. Trabecular bone contains platelike structures in addition to rodlike trabeculae. In addition, trabeculae are not perfectly straight (as assumed in the model) but are somewhat curved and jagged. Nevertheless, cylinderlike objects (trabeculae) contribute a substantial portion of the scattering. Moreover, only those platelike structures oriented approximately perpendicular to the ultrasound propagation direction can measurably affect the frequency dependence of scattering. Finally, plates are comparatively rare in bones from older subjects, upon which the data from the studies mentioned above are based.

The potential effects of coherent scattering, multiple scattering, small measurement volumes (often only a few wavelengths deep), ultrasonic beam distortion (due to propagation through an inhomogeneous medium, bone), and phase cancellation compromise accuracy of measurements of frequency-dependent backscatter. Given these numerous complicating factors, it may not be possible to ascertain at present the extent to which any discrepancy between measurements and theoretical predictions is attributable to imperfect modeling as opposed to experimental uncertainty.

Anderson, M. E., Soo, S. C., and Trahey, G. E. (**1998**). "Microcalcifications as elastic scatterers under ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 925–934.

Ashman, R. B., and Rho, J. Y. (**1988**). "Elastic modulus of trabecular bone material," J. Biomech. **21**, 177–181.

Bauer, D. C., Glüer, C. C., Cauley, J. A., Vogt, T. M., Ensrud, K. E., Genant, H. K., and Black, D. M. (**1997**). "Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women," Arch. Intern. Med. **157**, 629–634.

Bouxsein, M. L., and Radloff, S. E. (**1997**). "Quantitative ultrasound of the calcaneus reflects the mechanical properties of calcaneal trabecular bone," J. Bone Miner. Res. **12**, 839–846.

Chaffai, S., Roberjot, V., Peyrin, F., Berger, G., and Laugier, P. (**2000**). "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results," J. Acoust. Soc. Am. **108**, 2403–2411.

Chappard, C., Laugier, P., Fournier, B., Roux, C., and Berger, G. (**1997**). "Assessment of the relationship between broadband ultrasound attenuation and bone mineral density at the calcaneus using BUA Imaging and DXA," Osteoporosis Int. **7**, 316–322.

Faran, J. J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418.

Giat, P., Chappard, C., Roux, C., Laugier, P., and Berger, G. (**1997**). "Preliminary clinical assessment of the backscatter coefficient in osteoporosis," in Proc. 22nd Int. Symp. Ultrason. Imag. and Tissue Char., Washington, DC, 16 June 1997 (abstract).

Glüer, C., Cummings, S. R., Bauer, D. C., Stone, K., Pressman, A., Mathur, A., and Genant, H. K. (**1996**). "Osteoporosis: Association of recent fractures with quantitative US findings," Radiology **199**, 725–732.

Gong, J. K., Arnold, J. S., and Cohn, S. H. (**1964**). "Composition of trabecular and cortical bone," Anat. Rec. **149**, 325–332.

Grenoble, D. E., Katz, J. L., Dunn, K. L., Gilmore, R. S., and Murty, K. L. (**1972**). "The elastic properties of hard tissues and apatites," J. Biomed. Mater. Res. **6**, 221–233.

Hans, D., Dargent-Molina, P., Schott, A. M., Sebert, J. L., Cormier, C., Kotzki, P. O., Delmas, P. D., Pouilles, J. M., Breart, G., and Meunier, P. J. (**1996**). "Ultrasonographic heel measurements to predict hip fracture in elderly women: the EPIDOS prospective study," Lancet **348**, 511–514.

Hausler, K. D., Rich, P. A., Smith, P. C., and Barry, E. B. (**1999**). "Relationships between static histomorphometry and ultrasound in the human calcaneus," Calcif. Tissue Int. **64**, 477–480.

Hoffmeister, B. K., Whitten, S. A., and Rho, J. Y. (**2000**). "Low megahertz ultrasonic properties of bovine cancellous bone," Bone (N.Y.) **26**, 635–642.

Hosokawa, A., and Otani, T. (**1997**). "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **23**, 405–418.

Lang, S. (**1970**). "Ultrasonic method for measuring elastic coefficients of bone and results on fresh and dried bovine bones," IEEE Trans. Biomed. Eng. **BME-17**, 101–105.

Langton, C. M., Palmer, S. B., and Porter, R. W. (**1984**). "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91.

Langton, C. M., Njeh, C. F., Hodgskinson, R., and Carrey, J. D. (**1996**). "Prediction of Mechanical Properties of the Human Calcaneus by Broadband Ultrasonic Attenuation," Bone (N.Y.) **18**, 495–503.

Laugier, P., Droin, P., Laval-Jeantet, A. M., and Berger, G. (**1997a**). "*In vitro* assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone (N.Y.) **20**, 157–165.

Laugier, P., Giat, C., Chappard, C., Roux, C., and Berger, G. (**1997b**). "Clinical assessment of the backscatter coefficient in osteoporosis," in Proc. 1997 IEEE Ultrason. Symp., October 1997, pp. 1104–1105.

Luo, G., Kaufman, J. J., Chabrera, A., Bianco, B., Kinney, J. H., Haupt, D., Ryaby, J. T., and Siffert, R. S. (**1999**). "Computational methods for ultrasonic bone assessment," Ultrasound Med. Biol. **25**, 823–830.

Nicholson, P. H. F., Strelitzki, R., Cleveland, R. O., and Bouxsein, M. L. (**2000**). "Scattering of ultrasound in cancellous bone: predictions from a theoretical model," J. Biomech. **33**, 503–506.

Rho, J. Y., Ashman, R. B., and Turner, C. H. (**1993**). "Young's modulus of

trabecular and cortical bone material: ultrasonic and microtensile measurements,'' J. Biomech. **26**, 111–119.
Roberjot, V., Laugier, P., Droin, P., Giat, P., and Berger, G. (**1996**). ''Measurement of integrated backscatter coefficient of trabecular bone,'' Proc. 1996 IEEE Ultrason. Symp., Vol. 2, pp. 1123–1126.
Rossman, P., Zagzebski, J., Mesina, C., Sorenson, J., and Mazess, R. (**1989**). ''Comparison of Speed of Sound and Ultrasound Attenuation in the Os Calcis to Bone Density of the Radius, Femur and Lumbar Spine,'' Clin. Phys. Physiol. Meas. **10**, 353–360.
Roux, C., Roberjot, V., Paorcher, R., Kolta, S., Dougados, M., and Laugier, P. (**2001**). ''Ultrasonic backscatter and transmission parameters at the os calcis in postmenopausal osteoporosis,'' J. Bone Miner. **16**, 1353–1362.
Wear, K. A. (**1999**). ''Frequency dependence of ultrasonic backscatter from human trabecular bone: theory and experiment,'' J. Acoust. Soc. Am. **106**, 3659–3664.
Wear, K. A., and Armstrong, D. W. (**2000**). ''The relationship between ultrasonic backscatter and bone mineral density in human calcaneus,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 777–780.
Wear, K. A., and Armstrong, D. W. (**2001**). ''Relationships among calcaneal backscatter, attenuation, sound speed, hip bone mineral density, and age in normal adult women,'' J. Acoust. Soc. Am. **110**, 573–578.
Wear, K. A., and Garra, B. S. (**1997**). ''Assessment of bone density using broadband ultrasonic backscatter,'' in Proc. 22nd Int. Symp. Ultrason. Imag. and Tissue Char., Washington, DC, 14 (Abstract), June 1997.
Wear, K. A., and Garra, B. S. (**1998**). ''Assessment of bone density using ultrasonic backscatter,'' Ultrasound Med. Biol. **24**, 689–695.
Williams, J. L. (**1992**). ''Ultrasonic wave propagation in cancellous and cortical bone: predictions of some experimental results by Biot's theory,'' J. Acoust. Soc. Am. **92**, 1106–1112.
Zagzebski, J. A., Rossman, P. J., Mesina, C., Mazess, R. B., and Madsen, E. L. (**1991**). ''Ultrasound transmission measurements through the os calcis,'' Calcif. Tissue Int. **49**, 107–111.

Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations (L)

Douglas P. Nowacek
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 and Sensory Biology and Behavior Program, Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, Florida 34236

Brandon M. Casper
College of Marine Science, University of South Florida, 140 Seventh Avenue South, St. Petersburg, Florida 33701

Randall S. Wells and Stephanie M. Nowacek
Chicago Zoological Society, c/o Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, Florida 34236

David A. Mann[a)]
College of Marine Science, University of South Florida, 140 Seventh Avenue South, St. Petersburg, Florida 33701 and Sensory Biology and Behavior Program, Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, Florida 34236

(Received 25 October 2003; revised 11 April 2003; accepted 14 April 2003)

Recordings of manatee (*Trichechus manatus spp.*) vocalizations were made in Florida and Belize to quantify both intraspecific and geographic variation. Manatee vocalizations were relatively stereotypical in that they were short tonal harmonic complexes with small frequency modulations at the beginning and end. Vocalizations ranged from almost pure tones to broader-band tones that had a raspy quality. The loudest frequency was typically the second or third harmonic, with average received levels of the peak frequency of about 100 dB *re* 1 μPa. Signal parameters measured from these calls showed the manatees from Belize and Florida have overlapping distributions of sound duration, peak frequency, harmonic spacing, and signal intensity, indicating no obvious distinguishing characteristics between these isolated populations. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582862]

PACS numbers: 43.80.Ka, 43.30.Sf, 43.80.Ev [WA]

I. INTRODUCTION

Florida manatees, *Trichechus manatus latirostris*, are endangered, and many animals are killed or injured each year by boat strikes. At least 25% of annual documented manatee deaths are due to collisions with vessels (Marine Mammal Commission, 2002). Despite the fact that behavioral evidence indicates that manatees do have the ability to detect and respond to approaching vessels (Nowacek *et al.*, 2000; Weigle *et al.*, 1994), the animals continue to be hit. In an effort to contribute to solutions that might reduce manatee morbidity and mortality from boat strikes, we set out to build a device to alert boaters of the presence of manatee(s) based on passive detection of their sounds.

Most of the boat strike mitigation effort has been directed toward testing the hearing capabilities of manatees (Gerstein *et al.*, 1999; Ketten *et al.*, 1992) and their behavioral response in the presence of boats (Nowacek *et al.*, 2000; Weigle *et al.*, 1994) in hopes of understanding if and how they respond to oncoming vessels. Manatee sound production has been documented in only four papers (Evans and Herald, 1970; Schevill and Watkins, 1965; Sonoda and Takemura, 1973; Sousa-Lima *et al.*, 2002), including very little discussion of variability inherent in the animals' vocal repertoire. These reports describe several types of sounds, but the primary vocalization recorded was a tonal sound often having several harmonics with the second or third harmonic often stronger than the fundamental frequency; intraspecific variation was not reported. The fundamental frequency ranged from 2.5 to 5 kHz for the West Indian manatee, *Trichechus manatus spp.*, and 2.6 to 5.9 kHz for the Amazonian manatee, *Trichechus inunguis*.

Vocalizations may be the most easily detected sounds produced by manatees (compared to chewing and flatulence) because published reports show them to have the highest signal-to-noise ratio and the vocalizations occur in frequency bands most dissimilar to the natural background noise of the manatees' environment. To determine the amount of stereotypy in vocalizations we recorded sounds from Antillean manatees (*Trichechus manatus manatus*) in Southern Lagoon, Belize (Lat/Lon: 17° 12′N, 88° 20′W) and Florida manatees in Crystal River, FL (Lat/Lon: 28° 53′ N, 82° 35′W). These two groups of manatees are subspecies of the West Indian manatee, so comparing the structure of their vocalizations could add to our understanding of the similarities and differences between the two subspecies.

[a)]Author to whom correspondence should be addressed. Electronic mail: dmann@marine.usf.edu

 0001-4966/2003/114(1)/66/4/$19.00

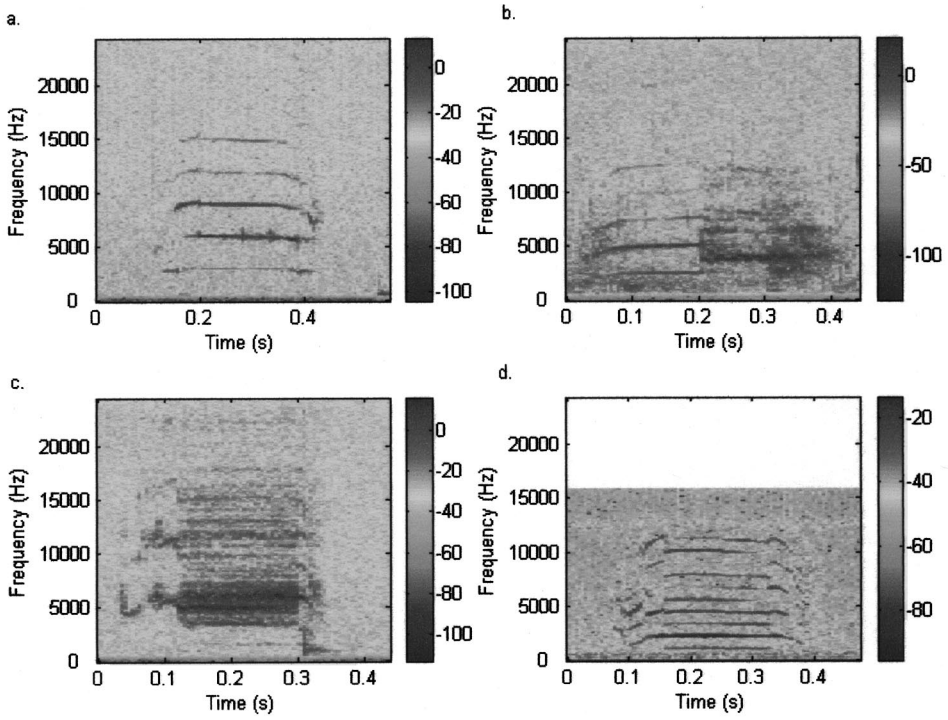


FIG. 1. Spectrograms of representative manatee vocalizations. (a) Typical tonal harmonic vocalization (Crystal River, FL), (b) tonal vocalization transitioning to less tonal (Crystal River, FL), (c) broader-band harmonic vocalization (Crystal River, FL), and (d) typical tonal harmonic vocalization (Southern Lagoon, Belize). The acoustic tag used to record manatees in Belize sampled at $F's = 32$ kHz. The scale bar shows relative sound levels in decibels.

II. METHODS

A. Belize recordings

In Belize we recorded sounds on a digital archival tag, the DTAG (Johnson and Tyack, 2003), which simultaneously records the attitude (pitch and roll), heading, and depth of the animal with sounds produced or received by the animal. The hydrophone signal was digitized continuously at 32 kHz. Selected manatees were encircled by a seine net in shallow water and brought aboard the capture vessel for health assessment, measurements, and tagging. DTAGs were attached to three individual manatees via temporary attachments using standard manatee peduncle belts. Floats tethered to the belts carried VHF transmitters for direct radio-tracking, and in some cases a GPS receiver. Sonic tags incorporated into the belts permitted underwater tracking. The tags were recovered either when the belts released via a corrodible link or animals were recaptured for recovery of the DTAG and removal of other tags and belts. Attachments lasted approximately 24 h with acoustic data recorded for 4 or 8 h depending on the tag settings used.

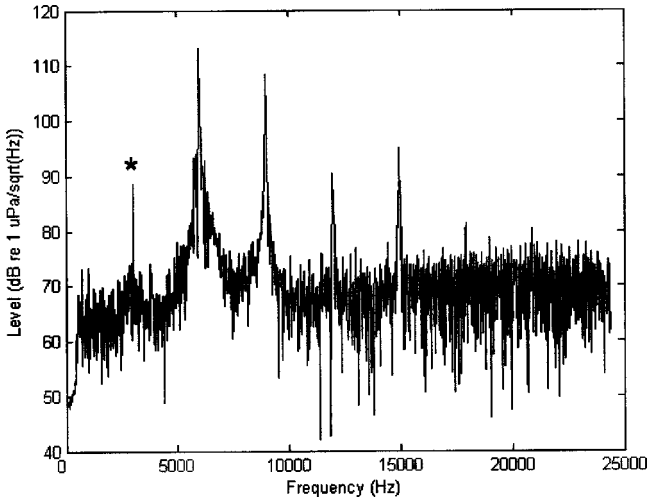


FIG. 2. Power spectrum of a typical manatee vocalization. Plot shows a 2048-point FFT of a 2048-point segment from the middle of the signal shown in Fig. 1(a). * indicates the fundamental frequency.

B. Crystal River, FL recordings

Recordings in Crystal River, FL were made using a hydrophone (HTI 96-min; sensitivity −164 dBV/μPa, 20 Hz to 32 kHz) connected to a data acquisition system (Tucker-Davis Technologies, RP2.1 24-bit ADC) and laptop computer. Signals were acquired continuously for approximately 2 h at 48 828.125 Hz sample rate from an anchored boat with the hydrophone on the bottom at 1-m depth. Recordings were made approximately 20 m away from a group of about 50 manatees that were in the spring. Individuals would occasionally move out of the spring into the spring run to within 2 m of the boat.

C. Data analysis

Data were analyzed with MATLAB (Mathworks, Inc.). Peak frequency (frequency with the most energy) and the level of the peak frequency were determined by performing a 1024-point FFT on a 1024-point segment of the middle of each call. This minimized smearing of frequencies due to changes in frequencies during the call. Harmonic spacing (which is equivalent to the fundamental frequency) was determined by performing an autocorrelation of the FFT, and measuring the first peak after lag zero. Duration was measured by MATLAB with an automatic detection algorithm, and then verified by hand from the spectrogram. While this may lead to a small uncertainty in measuring the time, it was advantageous for signals with low signal-to-noise ratios.

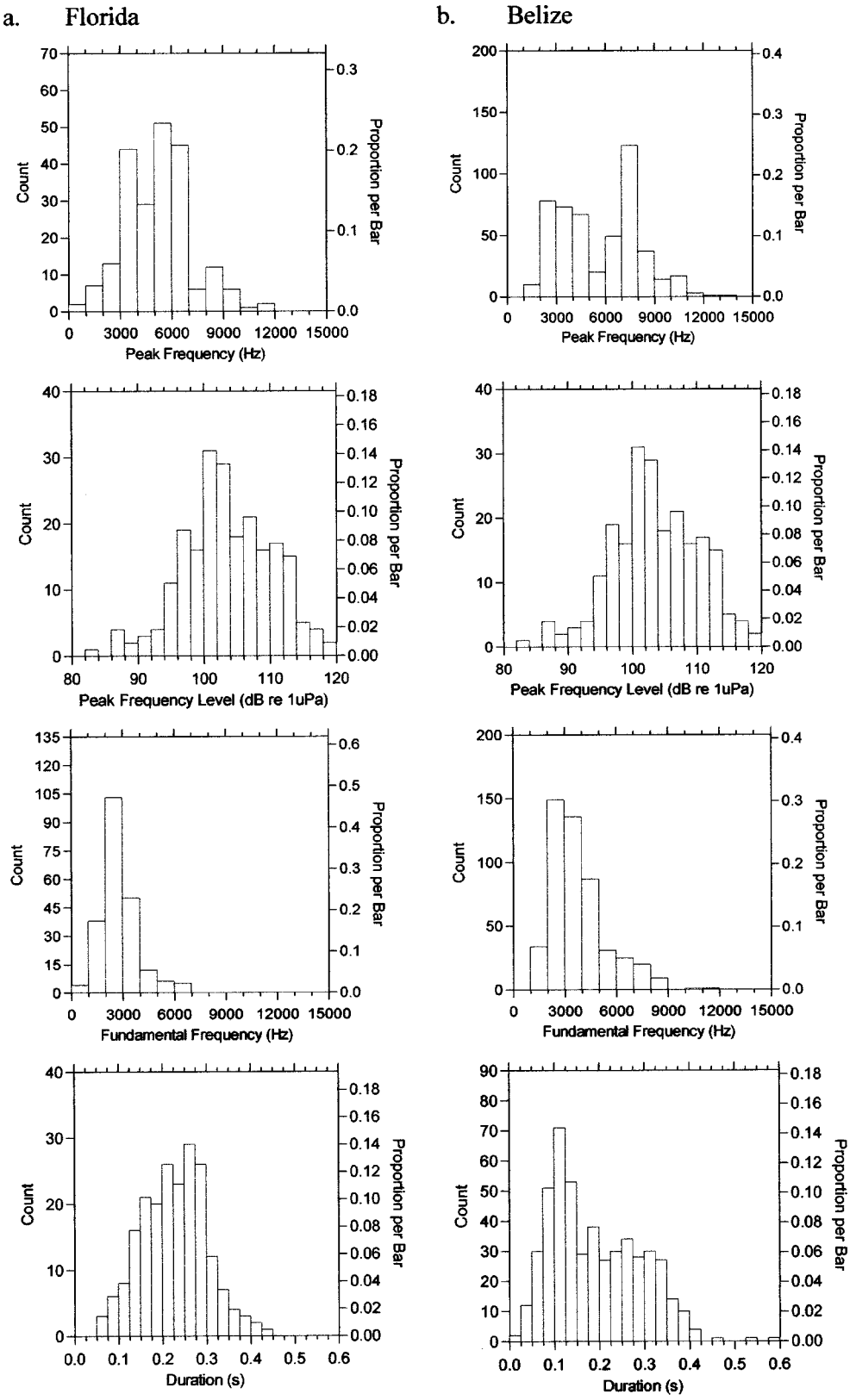


FIG. 3. Distribution of manatee vocalization parameters from (a) Crystal River, FL and (b) Southern Lagoon, Belize (data from 11 March 2002). Panels show peak frequency, level of peak frequency, fundamental frequency (harmonic spacing), and duration from top to bottom. We show only the Belize data from 11 March because this animal was in a similar behavioral situation to those in Crystal River. The animal was with a group of manatees, so these samples include sounds from other manatees as well as from the tagged animal. We felt that comparing sounds produced in similar situations provided the most accurate comparison for our study.

III. RESULTS

Manatees in Florida and Belize produced vocalizations that were harmonic complexes with small frequency modulations at the beginning and end (Fig. 1). These ranged from almost pure tones to broader-band tones that have a noisy quality. The loudest frequency was typically the second or third harmonic (Figs. 2 and 3). Signal parameters measured from these calls show the manatees from Belize and Florida have overlapping distributions of sound duration, peak frequency, harmonic spacing, and signal intensity (Table I and Fig. 3). Statistical comparisons between these signals were not possible because it is not known which individuals pro-

TABLE I. Basic statistics of sounds recorded from manatees in Southern Lagoon, Belize and Crystal River, FL. The mean for each data set is shown with standard deviation in parentheses.

	Belize 9 March 2002	Belize 10 March 2002	Belize 11 March 2002	Crystal River
n	26	105	208	218
Peak frequency (Hz)	3180 (727)	7080 (2207)	5560 (2559)	5223 (1937)
Peak level (dB *re* 1 μPa)	97.1 (4.3)	92.5 (6.6)	100.0 (4.7)	103.6 (6.8)
Fundamental frequency (Hz)	3180 (728)	4380 (1618)	3630 (1620)	2867 (1059)
Duration (s)	0.032 (0.017)	0.161 (0.10)	0.217 (0.098)	0.228 (0.074)

duced which sounds. Treating each call as a separate replicate would lead to pseudoreplication. The rate of vocalization in Crystal River, accounting for the number of animals present, was 1.29 vocalizations per minute, and in Belize it ranged from 0.09 to 0.75 per minute for the three animals tagged. When alone, the Belize animals were often silent for periods of $>$10 min. The rates of vocalization we measured were similar to earlier reports (Bengtson and Fitzgerald, 1985).

IV. DISCUSSION

There are no obvious differences in the vocalizations produced by manatees from Florida and Belize. For the parameters that were characterized (sound duration, peak frequency, and harmonic spacing) manatees from Florida and Belize had overlapping distributions. It is possible, however, that there are differences that we did not characterize.

The finding that the second or third harmonic of the vocalization is usually most intense could be due either to how the sound is produced, or to propagation effects where the lower frequencies do not propagate as well in shallow water, or a combination of the two (Rogers and Cox, 1988). Given that previous research on captive animals also found the fundamental to be less intense (Evans and Herald, 1970; Schevill and Watkins, 1965), our results are probably best explained by the production system of the animals. Indeed, in all of these recordings we do not know the distance to the sound-producing manatee, only a range to the manatees that were in the area. The data from Belize include both the animal wearing the DTAG as well as animals vocalizing nearby. Still the received levels can be taken to show the range of levels that might be produced by manatees, and they are likely within 6–15 dB of source levels given the range of distances over which the recordings were made (assuming losses of 3 dB per doubling of distance).

The motivation for this study was to determine the range of natural variability in manatee sounds, so that a passive acoustic detection device can be developed to warn boaters of the presence of manatees. These data show that West Indian manatee sounds are relatively stereotypical, even between subspecies, in that they are short tonal harmonic complexes, and lend themselves readily to this application.

ACKNOWLEDGMENTS

We thank John E. Reynolds III, Buddy Powell, Robert Bonde, Mesha Gough, and Kevin Andrewyn for their assistance in the field and/or their consultation on this manuscript and Mark Johnson and Alex Shorter for their assistance preparing and deploying the DTAGs. This project was supported by the Florida Fish and Wildlife Conservation Commission, Wildlife Trust, the Chicago Board of Trade, and the Chicago Zoological Society.

Bengtson, J. L., and Fitzgerald, S. M. (**1985**). ''Potential role of vocalizations in West Indian manatees,'' J. Mammal. **66**(4), 816–819.

Evans, W. E., and Herald, E. S. (**1970**). ''Underwater calls of a captive Amazon manatee, *Trichechus inunguis*,'' J. Mammal. **51**, 820–823.

Gerstein, E. R., Gerstein, L., Forsythe, S. E., and Blue, J. E. (**1999**). ''The underwater audiogram of the West Indian manatee (*Trichechus manatus*),'' J. Acoust. Soc. Am. **105**, 3575–3583.

Johnson, M. P., and Tyack, P. L. (**2003**). ''A digital acoustic recording tag for measuring the response of wild marine mammals to sound,'' IEEE J. Ocean. Eng. **28**(1), 3–12.

Ketten, D. R., Odell, D. K., and Domning, D. P. (**1992**). *Marine Mammal Sensory Systems*, edited by J. Thomas (Plenum, New York), pp. 77–95.

Nowacek, S. M., Wells, R. S., Nowacek, D. P., Owen, E. C. G., Speakman, T. R., and Flamm, R. O. (**2000**). ''Manatee behavioral responses to vessel approaches,'' Report No. FWC-00127, 2000.

Rogers, P. H., and Cox, M. (**1988**). *Sensory Biology of Aquatic Animals*, edited by J. Atema, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).

Schevill, W. E., and Watkins, W. A. (**1965**). ''Underwater calls of *Trichechus* (Manatee),'' Nature (London) **205**(4969), 373–374.

Sonoda, S., and Takemura, A. (**1973**). ''Underwater sounds of the manatees, *Trichechus manatus* and *T. inunguis* (Trichechidae),'' Report of the Institute for Breeding Research, Tokyo University of Agriculture, Vol. 4, pp. 19–24.

Sousa-Lima, R. S., Paglia, A. P., and Da Fonseca, G. A. B. (**2002**). ''Signature information and individual recognition in the isolation calls of Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia),'' Anim. Behav. **63**, 301–310.

U. S. Marine Mammal Commission (**2002**). ''Annual Report of the U.S. Marine Mammal Commission,'' U.S. Marine Mammal Commission, 4340 East West Highway, Suite 905, Bethesda, MD.

Weigle, B. L., Wright, I. E., and Huff, J. A. (**1994**). ''Responses of manatees to an approaching boat: a pilot study,'' presented at the First international manatee and dugong research conference, Gainesville, FL.

Estimation of the Nakagami parameter from log-compressed ultrasonic backscattered envelopes (L)

P. M. Shankar[a)]
Department of Electrical and Computer Engineering, Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104

(Received 20 December 2002; accepted for publication 21 April 2003)

In commercial medical ultrasonic imaging systems, B mode images are displayed with the log-compressed backscattered envelope. While the Nakagami distribution can describe the statistics of the backscattered envelope, this nonlinear compression of the envelope modifies the statistics of the image pixels from the Nakagami distribution. An estimation of the Nakagami parameter from the log-compressed data has been undertaken, enabling its application in tissue classification. Results based on random number and A-line simulations indicate that the Nakagami parameter m can be easily obtained from the log-compressed data. The values of the parameter estimated from the log-compressed data are very close to those obtained directly from the envelope. This suggests the use of Nakagami statistics based techniques for tissue characterization even when rf data are not available. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1581281]

PACS numbers: 43.80.Qf, 43.80.Vj, 43.80.Cs [FD]

I. INTRODUCTION

The Nakagami distribution is a reasonably good probability density function (pdf) for modeling statistics of the envelope of the backscattered echo from tissue in medical ultrasound.[1,2] Its parameters have the potential to discriminate benign and malignant masses in breast tissue.[3] The use of the Nakagami distribution and its parameters requires that the imaging systems provide access either to the rf data or envelope values. The present day medical imaging systems, however, use logarithmic compression to control the dynamic range of the image displays.[4–6] The density function of the log-compressed data will not be Nakagami distributed. This motivates the need to develop techniques that can estimate the Nakagami parameter from the log-compressed data, thereby enabling tissue characterization. The work described here explores estimation of the Nakagami parameter from the log-compressed data. Results on random number and A-line simulations are described.

II. THEORY

The probability density function of the Nakagami distributed envelope R is given by[1–3]

$$f(r)=\frac{2m^m r^{2m-1}}{\Gamma(m)\Omega^m}e^{-(m/\Omega)r^2}, \quad r\geqslant 0, \tag{1}$$

where m is the Nakagami parameter that determines the shape of the density function and Ω is the scaling factor given by

$$\Omega=E[R^2]. \tag{2}$$

The parameter m can be estimated as

$$m=\frac{\Omega^2}{E[R^4]-\Omega^2}. \tag{3}$$

In Eqs. (2) and (3), $E[.]$ stands for the statistical average. The Nakagami parameter takes values in the range $0<m<\infty$, with $m=1$ corresponding to Rayleigh statistics. It was shown to be capable of discriminating benign and malignant masses in B scan breast images.[3] This parameter was computed first by quadrature demodulating the rf data to get the in-phase and quadrature components. These components are subsequently used to get the envelope R. For display purposes, the envelope (or the intensity) values can be used directly. However, the dynamic range of these values is larger than the dynamic range of the display device, making displays based purely on the values of envelopes or intensities very difficult. Some form of nonlinear compression of the envelope values needs to be performed to reduce the range of values of the input to the display device. Most of the imaging systems use a logarithmic compression of the envelope to limit[4–6] the range of values and thereby control the dynamic range. The simplest logarithmic compression is the decibel converter. The input to the display device Z can be expressed as

$$Z=D\ \log_{10}R+G, \tag{4}$$

where D is the dynamic range parameter and G is the linear gain parameter. If $D=20$ and $G=0$, Eq. (4) becomes the simple decibel converter. Note that G does not affect the nature of the pdf of Z because it is an offset and will only lead to a different average value of Z. It will therefore be safe to assume that $G=0$. The effect of $G\neq 0$ is discussed later in this section and in the concluding remarks.

Given that R follows Nakagami statistics according to Eq. (1), the density function of Z can be derived using the transformation of random variables.[7] However, the interest does not lie in the pdf of Z, but in finding ways to estimate

a)Electronic mail: pshankar@coe.drexel.edu

 0001-4966/2003/114(1)/70/3/$19.00

the Nakagami parameter. This can be done using the moments of the random variable. We will rewrite Eq. (4) taking G to be zero and defining Z_0 as

$$Z_0 = \frac{D}{2}\log_{10} Y, \tag{5}$$

where $Y = R^2$. The density function $f_Y(y)$ of Y is the gamma density function[1,2]

$$f_Y(y) = \left(\frac{m}{\Omega}\right)^m \frac{y^{m-1}}{\Gamma(m)} e^{-(m/\Omega)y}, \quad y \geqslant 0. \tag{6}$$

The mean of Z_0 can be written as[8–10]

$$\langle Z_0 \rangle = \int_0^\infty \left[\frac{D}{2}\right] \log_{10} y \left(\frac{m}{\Omega}\right)^m \frac{y^{m-1}}{\Gamma(m)} e^{-(m/\Omega)y}\, dy. \tag{7}$$

Equation (7) can be simplified [Eq. (19) in Ref. 1, Eq. (4-352.1) in Ref. 8] to

$$\langle Z_0 \rangle = \left[\frac{D}{2}\right]\left[\frac{\psi(m) - \log_e(m/\Omega)}{\log_e 10}\right], \tag{8}$$

where $\psi(m)$ is the digamma function.[8] The second moment can be written as

$$\langle Z_0^2 \rangle = \int_0^\infty \left[\frac{D}{2}\right]^2 (\log_{10} y)^2 \left(\frac{m}{\Omega}\right)^m \frac{y^{m-1}}{\Gamma(m)} e^{-(m/\Omega)y} dy. \tag{9}$$

Equation (9) can be solved to lead to [Eq. (19) in Ref. 1, Eq. (4-358.2) in Ref. 8]

$$\langle Z_0^2 \rangle = \left[\frac{D}{2\log_e 10}\right]^2 \left\{\left[\psi(m) - \log_e\left(\frac{m}{\Omega}\right)\right]^2 + \psi'(m)\right\}, \tag{10}$$

where $\psi'(m)$ is the trigamma function.[8] From Eqs. (9) and (10), the variance of Z_0 becomes

$$\mathrm{var}(Z_0) = \langle Z_0^2 \rangle - \langle Z_0 \rangle^2 = \left[\frac{D}{2\log_e 10}\right]^2 \psi'(m). \tag{11}$$

Equation (11) for a simple decibel converter ($D=20$) becomes

$$\mathrm{var}(Z_0) = \left[\frac{10}{\log_e 10}\right]^2 \psi'(m) = 18.86\, \psi'(m). \tag{12}$$

Thus, the Nakagami parameter m can be estimated by solving the equation for variance given in Eq. (12). It requires the knowledge of the dynamic range parameter D which usually appears on the screen displaying the B mode image.[4,5] Since variance does not depend on the mean, variance of Z in Eq. (4) and variance of Z_0 in Eq. (5) will be identical and it is given in Eq. (12). Thus, the value of G will have no effect on the computation of m and information on G will not be necessary.

III. RESULTS AND DISCUSSION

Random number simulations were done to test whether the parameter m can be effectively estimated from Eq. (12). A set of (3000) Nakagami random numbers was generated in MATLAB®.[11] These random numbers were obtained from the gamma random number generator as discussed in previous research.[2] They were converted into decibel units as in Eq. (5) with $D=20$. For this set, the parameter m was estimated ($m_{\log}$) by solving Eq. (12) using the Optimization Toolbox (fsolve) in MATLAB. The Nakagami parameter (m_{dir}) was also estimated directly from the random numbers using Eq. (3). The simulations were repeated ten times and Table I contains the results. The mean of the estimates and the standard deviation (quantity in brackets) show that the log-compressed data can provide very good estimates of the Nakagami parameter. By choosing values of $G \neq 0$, there were no changes seen in the estimates as expected because the variance does not depend on the mean.

TABLE I. Results of random number simulations are tabulated. The input values of the Nakagami parameter (m), mean values (m_{dir} obtained from the envelopes and $m_{\log}$ obtained from the log-compressed values) of the estimates of the parameter and the standard deviations (in brackets) are shown.

m	m_{dir}	$m_{\log}$
0.4	0.40(0.01)	0.401(0.013)
0.6	0.599(0.02)	0.605(0.02)
0.8	0.796(0.017)	0.793(0.018)
0.95	0.954(0.025)	0.947(0.027)
1.2	1.216(0.03)	1.22(0.035)

A-line simulations were also undertaken to see how the estimates of m matched. A one-dimensional discrete scattering model was used in our simulation. Details of this simulation are available elsewhere.[2] An A-line of 3.7 cm was chosen, with the location of the scatterers being uniformly distributed in the interval {0, 3.7 cm}. The transmitted pulse was modeled as

$$P(t) = -t\, \exp(-4\beta^2 t^2)\sin(2\pi f_0 t), \tag{13}$$

where f_0 is the center frequency (3.5 MHz) and β (0.8 MHz) is the bandwidth. The backscattered signal, $s(t)$, is a superposition of echoes from all the scatterers present,

$$s(t) = \sum_{n=1} \alpha_n P\left(t - 2\frac{x_n}{c}\right), \tag{14}$$

where α_n is the amplitude of the nth scatterer assumed to have a gamma distribution, x_n is the location of the nth scatterer from the origin and c is the velocity of sound (1446 m/s). Simulations were carried out with three different number densities ($N_d = 3$, 6, and 15 scat/mm) of randomly located scatterers. The signal-to-noise ratio of the scattering amplitudes, SNR_α,

$$\mathrm{SNR}_\alpha = \frac{\langle \alpha \rangle}{\sqrt{\langle \alpha^2 \rangle - \langle \alpha \rangle^2}}, \tag{15}$$

was also varied (0.6, 1.2, and 2). The envelope values were log-compressed as in Eq. (5) with $D=20$. For each number density and signal-to-noise ratio of the scattering cross sections, simulations were done ten times to get the mean and the standard deviation (within brackets). The results of these simulations are given in Table II. It is seen that the two estimates of the Nakagami parameter, one obtained directly from the envelope (m_{dir}) and the other one obtained from the log-compressed envelope ($m_{\log}$), are reasonably close.

TABLE II. Results of A line simulations are tabulated. The input values of the number density (N_d/mm) of scatterers and signal-to-noise ratio of the cross sections (SNR_α) are shown. The mean values (m_{dir} obtained from the envelopes and $m_{\log}$ obtained from the log-compressed values) of the estimates of the Nakagami parameter and the standard deviations (in brackets) are shown.

N_d	SNR_α	m_{dir}	$m_{\log}$
3	0.6	0.21 (0.0055)	0.245(0.013)
3	1.2	0.37(0.015)	0.358(0.024)
3	2	0.45(0.02)	0.39(0.03)
6	0.6	0.354(0.018)	0.43(0.024)
6	1.2	0.6(0.016)	0.645(0.025)
6	2	0.74(0.024)	0.77(0.034)
15	0.6	0.57(0.03)	0.67(0.021)
15	1.2	0.83(0.03)	0.88(0.03)
15	2	0.91(0.02)	0.94(0.02)

IV. CONCLUDING REMARKS

The results show that the Nakagami parameter can be estimated from the log-compressed data by solving Eq. (12). The estimates are close enough to the values obtained directly from the envelopes. One may be tempted to ask why the envelope R should not be calculated if the dynamic range parameter D is known and then use the standard way of estimating m. The answer lies in the fact that the approach described in this work does not need the value of the linear gain factor G, which is required to calculate the envelope values prior to compression. Because we are dealing with the variance of the compressed data, the linear gain factor will have no effect on the estimation of m using Eq. (12). Thus, even when rf data may not be available, the Nakagami parameters may be estimated, making it possible to undertake classification of tissue in ultrasonic B scans.

ACKNOWLEDGMENT

The work was supported by Grant No. NIH-NCI #P01-52823.

[1] M. Nakagami, ''The m-distribution—a general formula of intensity distribution in rapid fading,'' in *Statistical Methods on Radio Wave Propagation*, edited by W. C. Hoffman (Pergamon, New York, 1960), pp. 3–36.
[2] P. M. Shankar, ''A general statistical model for ultrasonic scattering from tissues,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**(3), 727–736 (2000).
[3] P. M. Shankar, V. A. Dumane, J. M. Reid, V. Genis, F. Forsberg, C. W. Piccoli, and B. B. Goldberg, ''Classification of ultrasonic B mode images of breast masses using the Nakagami distribution,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(2), 569–580 (2001).
[4] D. Kaplan and Q. Ma, ''On the statistical characteristics of log- compressed Rayleigh signals,'' J. Acoust. Soc. Am. **95**, 1396–1400 (1994).
[5] V. Dutt and J. F. Greenleaf, ''Statistics of the log-compressed echo envelope,'' J. Acoust. Soc. Am. **99**, 3817–3825 (1996).
[6] J. M. Sanches and J. M. Marques, ''3D reconstruction from log compressed Rayleigh images,'' in Proceedings of the 2001 International Conference on Image Processing (2001), Vol. 1, pp. 345–348.
[7] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. (McGraw–Hill, New York, 2002).
[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. (Academic, San Diego, 1994).
[9] S. Ghofrani, M. R. Jahed-Motlagh, and A. Ayatollahi, ''An adaptive speckle suppression filter based on Nakagami distribution,'' in Proceedings of EUROCON'2001, International Conference on Trends in Communications (2001), Vol. 1, pp. 84–87.
[10] Y-C. Ko and M. Alouni, ''Estimation of local mean power over Nakagami fading channels,'' in Proceedings of the 12th International Symposium on Personal, Indoor, and Mobile Radio Communications (2001), Vol. 1, pp. 107–112.
[11] MATLAB, Version 6.5, Release13 (MathWorks, Natick, MA, 2002).

Acoustic wave propagation in double porosity media

Xavier Olny[a] and Claude Boutin[b]
Départment Génie Civil et Bâtiment, URA CNRS 1652, Ecole Nationale des Travaux publics de l'Etat, Vaulx-en-Velin 69518 Cedex, France

(Received 29 March 2002; revised 12 October 2002; accepted 11 November 2002)

A study of the propagation of waves in porous media with an interconnected network of pores and micropores of very different characteristic sizes, saturated by a compressible Newtonian fluid, is proposed. With this aim, the homogenization technique for periodic separated scales media, is applied to realistic double porosity materials with motionless skeleton. From preliminary explicit estimations of wavelengths in the two fluid networks, it is shown that the macroscopic descriptions depend on the contrast of static permeability between pores and micropores and on frequency. The local equations are solved in the cases of low and high contrasts of permeability, and two main macroscopic behaviors are obtained. In the low contrast situation, the macroscopic flow is given by a kind of generalized Darcy's law involving both pores and micropores, and their respective characteristic frequencies. Regarding compressibility effects, both pore networks act in parallel. The high permeability contrast reveals that the macroscopic flow law is governed by the pores. The microporous part of the material is submitted to pressure diffusion effects, bringing dissipation, and modifying the dynamic bulk modulus of the material. The two situations of coupling are illustrated for simple geometry of double porosity materials, including perforated-and slits-microporous materials.
[DOI: 10.1121/1.1534607]

PACS numbers: 43.20.Bi, 43.20.Gp, 43.20.Jr, 43.55.Ev [MO]

I. INTRODUCTION

The description of wave propagation in porous media is driven by the need of understanding and improving performances of passive noise reducing devices. Basically, porous materials are biphasic, made of a solid skeleton saturated with a fluid. In most of the works dedicated to acoustics of porous media, a ''single'' porosity is assumed, because a characteristic microscopic size can be defined from the heterogeneities of the structure.[1–6]

In this paper, the porous medium under consideration is supposed to have a ''double'' porosity. This means that two interconnected networks of pores of very different characteristic sizes can be identified in the material. For instance, it can be depicted as a fractured material with a porous frame (geophysics), or as a porous medium with a microporous skeleton (acoustics).

Interest for such systems came at first from geophysics. The notion of double porosity, or double permeability is borne from studies carried out on naturally fractured porous rocks, such as oil fields. The benefits of describing oil flow and stock capacity in these kinds of soils justified theoretical studies undertaken during the 1960s. These works mainly led to phenomenological models, describing the quasipermanent flow of a weakly compressible fluid,[7,8] and showed the complexity of describing interscale couplings in the heterogeneous medium.

More recently, Auriault and Boutin resumed the study of deformable double porosity media using the homogenization method for periodic structures (HPS), introduced by Sanchez-Palencia.[9] They successively treated the cases of quasistatic regimes[10,11] and acoustics.[12] In their works, the solid material constituting the skeleton is assumed to be elastic, and the saturating fluid viscous, and incompressible. An important result pointed out by Auriault and Boutin is that the various macroscopic descriptions obtained depend on the ratios between the different characteristic sizes of the pores, the fractures, and the macroscopic medium. Following Auriault and Boutin, Royer[13] examined the problem of flow in coal–gas systems (compressible fluid), in the cases of an elastic and rigid (motionless) frame. Under this latter hypothesis, she exhibited three macroscopic behaviors corresponding two three interscales coupling situations between pores and fractures that go from a total coupling between the two networks to a no-coupling. The intermediate case is of particular interest because it points out that, under high compression levels, the pressure in porous element is nonlinearly dependent on the pressure in the fractures, and reveals a partial interscale coupling.

This case was thoroughly studied by Boutin *et al.*[14] in the context of acoustics. They showed that for small harmonic solicitations, pressures in the networks of pores and micropores are linearly related. Then, adding thermal effects, they proposed a general macroscopic model taking pressure diffusion effects in the microporous domain into account.

Either in Royer's, or Boutin *et al.*'s, theoretical works, frequency is implicitly assumed small enough so that inertial forces are neglected in micropores. A consequence of this hypothesis is that the flow in the micropores does not participate in the macroscopic flow.

Our purposes in this paper are to highlight interscale couplings that can exist in double porosity media, and to find

[a]Electronic mail: xavier.olny@entpe.fr
[b]Electronic mail: claude.boutin@entpe.fr

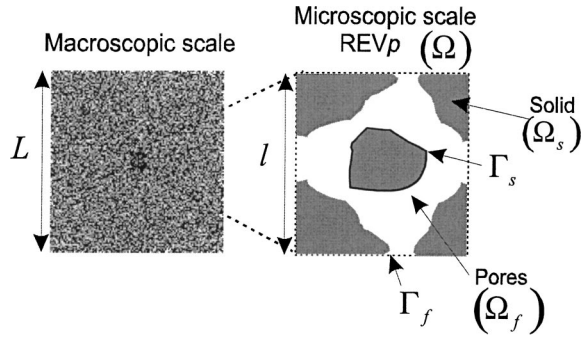


FIG. 1. Main notations for the single porosity medium from the macroscopic and microscopic points of view.

the corresponding macroscopic descriptions. With these aims, hypotheses are enlarged, taking into account the dynamic behavior of the microporous part of the material, in order to find the various macroscopic behaviors corresponding to ''realistic'' double porosity media.

The contributions presented in this paper are part of a Ph.D. thesis devoted to acoustics in double porosity media,[15] and including experimental validations.

In Sec. II, we present the physical analysis of double porosity media, which is the essential preliminary step before applying the asymptotic development technique. The heterogeneous medium is supposed to have a motionless solid skeleton and it is saturated by a compressible Newtonian fluid. A brief recall of sound propagation in ''single'' porosity media is presented, and differences with double porosity materials are clarified. Realistic double porosity materials are used to point out two coupling situations of particular interest, related to the contrast of static permeability between pores and micropores.

In Secs. III and IV, the HPS method is applied to the low and high permeability contrast situations. The so-obtained macroscopic descriptions are then discussed, and illustrated for simple geometries of materials with double porosity.

II. HOMOGENIZATION TECHNIQUE APPLIED TO DOUBLE POROSITY MEDIA

A. Single versus double porosity media

Before tackling the problem of acoustic wave propagation in double porosity media, the main results obtained in the case of single porosity have to be recalled.

Usual porous materials, such as reticulated foams, mineral wools, or granular materials, are characterized by the fact that it is generally possible to identify, from the structure, one representative elementary volume (REV) of characteristic size l, representative of the heterogeneities size (see Fig. 1).

Under acoustical solicitations, assuming that the wavelength remains large compared to l, the heterogeneous medium can be assimilated, from the macroscopic point of view (related to the wavelength), as a homogenous medium in which three waves can *a priori* propagate:[1,2] two compression waves and a shear wave. In this work, the solid phase is assumed motionless, and only the wave propagation in the fluid phase is considered.

Air in pores is taken as a compressible Newtonian viscous fluid. Under harmonic acoustical solicitations ($e^{j\omega t}$, ω pulsation, $j=\sqrt{-1}$), the dynamic flow in the homogeneous equivalent medium (the macroscopic point of view) is given by the generalized or dynamic Darcy law,[16]

$$\langle \boldsymbol{\nu} \rangle_\Omega = -\frac{[\Pi]}{\eta} \boldsymbol{\nabla} p, \tag{1}$$

relating the mean velocity ($\langle \boldsymbol{\nu} \rangle_\Omega$) in a REV($\Omega$) ($\langle \cdot \rangle_\Omega = (1/\Omega)\int_{\Omega_f} \cdot \, d\Omega$), to the macroscopic pressure gradient ($\boldsymbol{\nabla} p$); $\eta (\approx 1.84\times 10^{-5}$ USI) is the kinematic viscosity of air.

$[\Pi]$ is the intrinsic permeability tensor, giving the filtering behavior of the material. It is complex valued and depends on the frequency and on the microstructure of the material.

Regarding the flow, the dynamic behavior of single porosity media can be divided into two frequency ranges.

(i) At low frequency, viscous forces dominate. In a first approximation, the flow in the pores is govern by a static Darcy law, the intrinsic permeability being roughly constant. It can be estimated in function of the porosity ($\phi = \Omega_f/\Omega$), and the characteristic size of pores (l):

$$\Pi(0) = O(\phi l^2), \quad \text{for } \omega \ll \omega_\nu. \tag{2}$$

For example, in the case of slit pores of width $2a$, the intrinsic permeability of one pore is $a^2/3$, in the direction of slits.

(ii) At high frequency, viscous forces become small compared to inertial forces. Regarding the first-order approximation, Π is imaginary, and decreases as the inverse of frequency:

$$\Pi(\infty) \approx \eta\phi / j\omega\rho_0\alpha_\infty \quad \text{for } \omega \gg \omega_\nu. \tag{3}$$

Here, $\rho_0 (\approx 1.2$ kg/m$^3)$ is the density of air and α_∞ is the tortuosity of the medium. The cutoff frequency of this kind of low-pass filter, also called ''viscous characteristic frequency,'' can be estimated from the intersection of these two asymptotic behaviors:

$$\omega_\nu = O(\eta / l^2 \rho_0 \alpha_\infty). \tag{4}$$

Assuming that the solid phase remains isothermal, air in the pores is submitted to thermal exchanges. The macroscopic law governing the average acoustical, or excess temperature ($\langle \tau \rangle_\Omega$) in the equivalent homogenous medium can be written in a very similar way as the generalized Darcy law,[17]

$$\langle \tau \rangle_\Omega = j\omega \frac{\Theta}{\kappa} p, \tag{5}$$

$\kappa (\approx 0.026$ Wm^{-1} K$^{-1})$ is the thermal conductivity of air.

Here Θ, introduced by Lafarge *et al.* as the ''thermal permeability'' (and initially noted as k'), is a scalar complex function, depending on frequency and on the microstructure. The frequency dependence of Θ is very similar to the one of Π.

At low frequency, Θ tends to its static value, which can be estimated as

$$\theta(0)=O(\phi l^2), \quad \text{for } \omega \ll \omega_t. \tag{6}$$

At high frequency, we have

$$\Theta(\infty)\approx \kappa\phi/j\omega\rho_0 C_p, \quad \text{for } \omega \gg \omega_t; \tag{7}$$

$C_p(\approx 1000\ \mathrm{J\,K^{-1}\,kg^{-1}})$ is the specific heat of air at constant pressure.

ω_t is the thermal characteristic frequency of the pores, estimated as

$$\omega_t = O\left(\frac{\kappa}{l^2\rho_0 C_p}\right). \tag{8}$$

Note that $\Theta(0)$ and $\Pi(0)$, as well as ω_t and ω_ν, are generally of the same order of magnitude.

Introducing $\delta_t(=(\kappa/\rho_0 C_p\omega)^{1/2})$, the thermal skin depth, the temperature variations in the pores leads to a modification of the macroscopic bulk modulus K,

$$K=\rho_0\frac{p}{\langle\rho\rangle_\Omega}=\frac{\gamma P_0/\phi}{\gamma-j(\gamma-1)\dfrac{\Theta}{\delta_t^2\phi}}. \tag{9}$$

Here, $P_0(\approx 101\,325\ \mathrm{Pa})$ is the static pressure, $\gamma(\approx 1.4)$ is the ratio between specific heats, and ρ is the acoustical density of air in pores.

K varies from its isothermal value (P_0/ϕ) at low frequency to its adiabatic value $(\gamma P_0/\phi)$ at high frequency.

The macroscopic wave equation for the equivalent homogeneous medium, taking into account viscoinertial and thermal effects, is finally written as

$$\frac{j\omega}{K}p-\boldsymbol{\nabla}\cdot\left(\frac{[\Pi]}{\eta}\boldsymbol{\nabla}p\right)=0. \tag{10}$$

In the case of an isotropic medium $([\Pi]=\Pi.[I])$, the sound speed is written as

$$c=\sqrt{j\omega K\frac{\Pi}{\eta}}, \tag{11}$$

from which we deduce the wavelength $(\lambda=c/f)$ of the compression wave in the porous material. Let us consider a plane wave $[p=e^{-j(2\pi/\lambda)\cdot u}(e^{j\omega t})]$ propagating in the **u** direction of the medium. In this direction the pressure gradient is

$$\frac{\partial p}{\partial u}=-j\frac{2\pi}{\lambda}p. \tag{12}$$

This result helps us to estimate the macroscopic size L, characteristic of the phenomenon $[|\partial p/\partial u|=O(p/L)]$:

$$L=O\left(\left|\frac{\lambda}{2\pi}\right|\right). \tag{13}$$

Equation (13) tells us that L is frequency dependant.

The so-recalled results highlight the consequences of assuming a single porosity in the porous medium: macroscopic functions (Π, K, λ) directly depend on the characteristic size l of the REV.

Now, let us consider a double porosity medium. Three scales are needed to describe the structure (Fig. 2). Let L, l_p, l_m, be, respectively, the three characteristic sizes for each level of representation: macroscopic, mesoscopic, and microscopic. In concrete terms, l_m and l_p are, respectively, given by the order of magnitude of the micro- and mesoheterogeneities (or micropores and pores), whereas the wavelength in the material can be used to evaluate L.

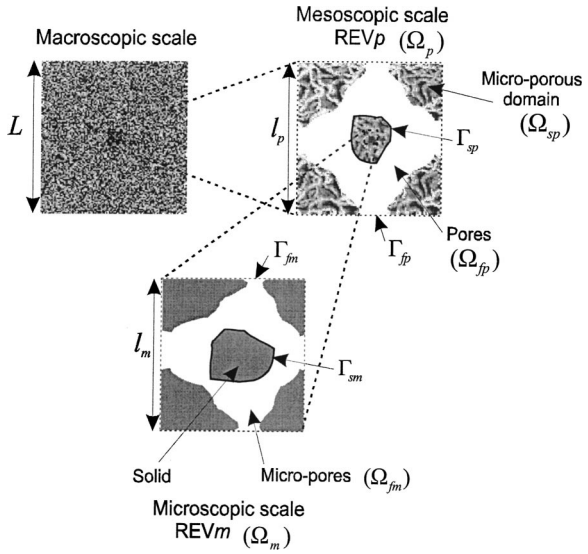


FIG. 2. The three scales of the double porosity medium.

In what follows, the subscripts *p, m, db* are used to distinguish variables, respectively, related to pores, micropores, and to the double porosity medium.

In this system, one assumes the existence of two representative elementary volumes REVm, and REVp defined, respectively, from the micropores and pores structures. Pores and micropores are supposed to be air-filled, and interconnected. The porosity of the double porosity medium can be written in function of porosities related to pores and micropores:

$$\phi_{\mathrm{db}}=\frac{\Omega_{\mathrm{fp}}}{\Omega_p}+\left(\frac{\Omega_{\mathrm{sp}}}{\Omega_p}\right)\frac{\Omega_{\mathrm{fm}}}{\Omega_m}=\phi_p+(1-\phi_p)\phi_m. \tag{14}$$

In order to represent the multiscale heterogeneous medium as a homogeneous equivalent one, the first assumption that has to be made is that the characteristic size of the phenomenon is large compared to the ''largest'' heterogeneities. This leads to an introduction of ε, expressing the first separation of scales (macro/meso):

$$\varepsilon=\frac{l_p}{L}\ll 1. \tag{15}$$

Note that ε is *a priori* frequency dependent [see Eq. (13)].

A second parameter ε_0 is now needed to quantify the separation of scales between micro- and mesoscales:

$$\varepsilon_0=\frac{l_m}{l_p}\ll 1. \tag{16}$$

As a result, permeability and bulk modulus related to pore and micropore structures, as well as the viscous and thermal characteristic frequencies, are shifted [see Eqs. (4) and (8)], and new macroscopic descriptions can appear depending on the value of ε_0.

B. Estimation of interscale couplings

In this part the consequences of the double separation of scales are studied. This analysis is supposed to lead to the various coupling behaviors. Physics in double porosity media are *a priori* noticeably different than in single porosity materials, but some results can be used to justify initial estimations in the triple scale material.

With this purpose, two ''fictive'' single porosity media are introduced.

The ''porous medium'' is defined from the mesostructure (REV Ω_p) replacing the microporous domain by an impervious skeleton.

The ''microporous medium'' is built from the microporous REVm (without pores). Classical acoustic parameters are associated to both fictive media such as permeability (Π_i), bulk modulus (K_i), porosity (ϕ_i), tortuosity ($\alpha_{\infty i}$) etc. ($i=m,p$).

Wavelengths (λ_i) in these two fictive single porosity media are now evaluated in order to estimate ranges of variation of parameters in pores and micropores. For the sake of simplicity, the asymptotic behaviors of the (Π_i), given in Eqs. (2) and (3), are used. Because of the small range of variations of the bulk modulus with frequency, K_i can be for example estimated by their adiabatic values ($\gamma P_0/\phi_i$), in pores and micropores. Since $\gamma \approx 1.4$, using the isothermal estimation wouldn't change the first-order estimations. Using Eqs. (2), (3), and (11), the low frequency estimation of the wavelength modulus ($|\lambda_i/2\pi|$, $i=m,p$), in the two fictive single porosity media, is given by

$$\left|\frac{\lambda_i}{2\pi}\right| = O\left(\frac{l_i}{\delta_\nu}\frac{\lambda_0}{2\pi}\right), \quad \omega \ll \omega_{\nu i}, \tag{17}$$

Regarding acoustical porous materials, common values for the tortuosity rarely exceed few units. That's why we'll consider in the following estimation that $\alpha_{\infty i}=O(1)$. For media with high tortuosity values, these estimations should be reconsidered. Thus, the high frequency estimation of the wavelength yields

$$\left|\frac{\lambda_i}{2\pi}\right| = O\left(\frac{1}{\omega}\sqrt{\frac{\gamma P_0}{\rho_0}}\right) = O(\lambda_0/2\pi), \quad \omega \gg \omega_{\nu i}, \tag{18}$$

introducing the viscous skin depth $\delta_\nu (=\sqrt{\eta/\rho_0\omega})$, and the wavelength in free air λ_0.

The separation of scales between pores and micropores induces a separation of the viscous characteristic frequencies. Assuming that the tortuosity of pores and micropores is $O(1)$, which is an acceptable hypothesis for most of the porous materials, the ratio between viscous characteristic frequencies in pores and micropores is given by

$$\frac{\omega_{\nu p}}{\omega_{\nu m}} = O\left(\frac{l_m^2}{l_p^2}\right) = O(\varepsilon_0^2). \tag{19}$$

Hence, three frequency domains appear corresponding to three situations of wave propagation in pores and micropores.

(i) At very low frequency ($\omega \ll \omega_{\nu p}$), in the viscous domain, waves are diffusive in both pores and micropores. The wavelengths in the two networks are smaller than λ_0. Moreover, the ratio between λ_m and λ_p is given by

$$\left|\frac{\lambda_m}{\lambda_p}\right| = O(\varepsilon_0). \tag{20}$$

(ii) In the middle-frequency range ($\omega_{\nu p} < \omega < \omega_{\nu m}$), the flow remains viscous in micropores, and is inertial in pores. The discrepancy between the wavelengths decreases with the viscous skin depth:

$$\left|\frac{\lambda_m}{\lambda_p}\right| = O\left(\frac{l_m}{\delta_\nu}\right). \tag{21}$$

(iii) At high frequency ($\omega \gg \omega_{\nu m}$), when δ_ν becomes much smaller than the micropores size, the inertial flow is reached in the whole fluid network, and the wavelength of the propagative waves in pores and micropores are of the same order of magnitude. They can be approximated by the wavelength in free air:

$$|\lambda_m| = O(|\lambda_p|) = O(\lambda_0). \tag{22}$$

From this point, several important remarks can be done.

Below $\omega_{\nu m}$, λ_p is much greater than λ_m, so that the macroscopic behavior of the double porosity medium should be imposed by physics in the pores. It means that λ_p can be used to estimate the wavelength in the double porosity medium, and then the macroscopic length L [see (13)]. As a result, at the mesoscale the boundary conditions on the microporous domain ($\Omega_{\rm sp}$) are quasihomogeneous. Therefore, the properties of $\Omega_{\rm sp}$ are the same as the one of the fictive microporous medium, and λ_m can be used to estimate the wavelength in $\Omega_{\rm sp}$.

These conclusions are still true above $\omega_{\nu m}$, simply because λ_p and λ_m are of the same order. Then, properties of the two fictive porous media can be used to estimate physics in the double porosity material.

The estimation of ε, is directly deduced from the previous remarks,

$$\varepsilon = \frac{l_p}{L} = O\left(2\pi\frac{l_p}{|\lambda_p|}\right). \tag{23}$$

This relation, of particular interest, will be used to estimate the relation between ε and ε_0, with respect to frequency.

C. Realistic materials and coupling situations

General estimations have been proposed. It is now necessary to consider realistic situations of triple scale materials to identify the various situations of couplings. The choice of the realistic double porosity media is based on three remarks.

(i) First, the mesoheterogeneities must be small enough so that they remain small compared to the wavelength in the audible frequency band ($\varepsilon \ll 1$). Then, we will suppose that $l_p \leqslant 10^{-2}$ m.

(ii) Second, we will consider that $l_m \geqslant 10^{-5}$ m, for the microporous medium to be pervious to acoustical waves.

(iii) Third, the double separation of scales imposes that the ratio between pores and micropores sizes is sufficient: $l_p/l_m > 10$.

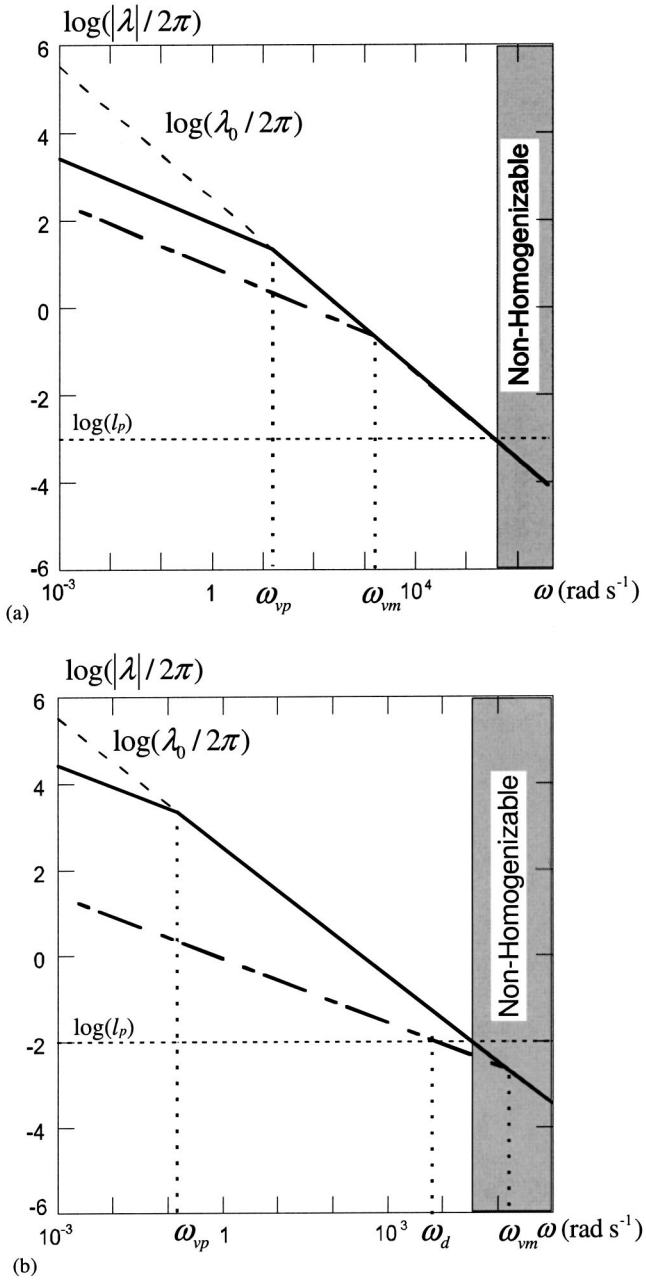


FIG. 3. A comparison between wavelengths, in the pores (thick line), in the micropores (broken line), and in free air (dashed line), in the case of a low (a) and high (b) permeability contrast.

According to these last limitations, attention is focused on two cases of main interest.

The first one refers to a relative low ratio between l_p and l_m, and then to a low static permeability contrast between the pores and micropores. Figure 3(a) depicts graphically, on a log–log plot, the estimations of the wavelengths calculated using Eqs. (17) and (18), with $l_p = 10^{-3}$ m and $l_m = 10^{-4}$ m. It appears that, in the homogenizable frequency range ($|\lambda_p/2\pi| \gg l_p$), $\omega_{\nu m}$ is small enough so that ($|\lambda_m/2\pi| \gg l_p$). Moreover, the inertial flow domain can be reached in both micropores and pores [$\omega = O(\omega_{\nu m})$], which means that, in this frequency range, a strong coupling between the two networks can be supposed.

The second identified situation is obtained with a relatively small value of ε_0. It corresponds to a high contrast of static permeability. Figure 3(b) illustrates such a case, considering extreme values for l_p and l_m (respectively, 10^{-2} and 10^{-5} m). In the homogenizable frequency range, l_m is small enough compared to l_p, so that a ''wavelength in the micropores'' remains smaller than in the pores. Furthermore, $|\lambda_m/2\pi|$ is of the same order of magnitude as l_p. This occurs around a characteristic frequency ω_d, estimated using Eq. (17):

$$\left|\frac{\lambda_m}{2\pi}\right| = O\left(\frac{l_m}{\sqrt{\eta/\rho_0\omega_d}}\frac{\lambda_0}{2\pi}\right) = l_p, \tag{24}$$

so that

$$\omega_d = O\left(\gamma P_0 \frac{l_m^2}{\eta}\frac{1}{l_p^2}\right). \tag{25}$$

It must be noticed that this phenomenon can only occur when waves in the microporous domain are of the diffusive kind. In this frequency range the ratio between wavelengths is

$$\left|\frac{\lambda_m}{\lambda_p}\right| = O(\varepsilon). \tag{26}$$

Considering that l_p is the characteristic size of the pores but also of the microporous domain, this means that around ω_d, pressure variations generated in the microporous domain *a priori* evolves at the mesoscopic scale [Eq. (12)]. The meaning of ω_d will appear clearly during the homogenization process.

D. Asymptotic development method

The HPS method, using the asymptotic expansion of unknowns, is a rigorous technique to obtain the macroscopic description of the medium from the description at local (microscopic and mesosopic) scales. The mathematical developments have been done using the formalism used by Auriault and Boutin to treat the problem of multiscale heterogeneous media.[10,18] With this aim, the double periodicity of the medium is assumed, and one elementary period at each scale of representation gives the representative elementary volume (see Fig. 2).

Three dimensionless space variables are needed to describe phenomena at each scale: let $\mathbf{X}$ be the space variable, and $\mathbf{x} = \mathbf{X}/L$, the dimensionless macroscopic space variable. In the same way, $\mathbf{y} = \mathbf{X}/l_p$ and $\mathbf{z} = \mathbf{X}/l_m$ are the dimensionless space variables for the meso- and microscale. The relations between these variables are simply given by $\mathbf{y} = \varepsilon^{-1}\mathbf{x}$ and $\mathbf{z} = \varepsilon_0^{-1}\mathbf{y}$.

Acoustical unknowns (pressures, velocities, densities, temperatures) will now be looked for in the form of functions depending on these three space variables. These functions are written in the micropores:

$$h_m = h_m(\mathbf{x}, \mathbf{y}, \mathbf{z}, \omega),$$

and, in the pores,

$$g_p = g_p(\mathbf{x}, \mathbf{y}, \omega),$$

where unknowns are $\mathbf{z}$ independent [except in the thin boundary layer of the microporous domain (Γ_{sp})].

As a consequence of the assumed double periodicity of the medium, the g_p functions are Ω_p periodic, and h_m functions are Ω_p periodic and Ω_m periodic.

The process consists of four steps.

(i) The first one, which is also the key point of the method, consists of estimating the relative weight of the terms in local equations with the aim of normalizing them. This will be based on the physical analysis presented in the preceding section (Sec. II B).

(ii) Second, asymptotic developments of acoustical functions (velocity, temperature, pressure, and density), are looked for in the following forms:

$$g_p=\sum_{i=0}^{\infty}\varepsilon^i g_p^i, \quad \text{in the pores},$$

$$h_m=\sum_{j=0}^{\infty}\varepsilon_0^j h_m^j, \quad \text{in the micropores}.$$

They are injected into the normalized equations, and the terms of the same order are identified.

(iii) Third, the so-obtained systems of differential equations are solved using variational formulations.

(iv) Fourth and last, the macroscopic dynamic behavior naturally appears expressing the compatibility between balance equations and the periodicity of local variables.

E. Local equations—Dimensionless numbers

1. Equations governing the flow

The case of small acoustic perturbations is considered. The local dynamic equation of motion, considering a Newtonian fluid, is given by the linearized Navier–Stokes equation:

$$\eta\,\Delta\boldsymbol{\nu}_i-\boldsymbol{\nabla}p_i=j\omega\rho_0\boldsymbol{\nu}_i, \quad i=m,p. \tag{27}$$

Regarding boundary conditions, the solid phase (skeleton) is supposed to be motionless,

$$\boldsymbol{\nu}_{m/\Gamma_{sm}}=\mathbf{0}, \tag{28}$$

and the continuity of displacements between pores and the microporous domain on the boundary Γ_{sp} yields

$$\boldsymbol{\nu}_{p/\Gamma_{sp}}=\langle\boldsymbol{\nu}_m\rangle_{\Omega_m}=\frac{1}{\Omega_m}\int_{\Omega_{fm}}\boldsymbol{\nu}_m\,d\Omega. \tag{29}$$

The continuity of pressure on the interface Γ_{sp} reads as

$$p_m=p_p, \quad \text{on } \Gamma_{sp}. \tag{30}$$

Dimensionless numbers. From Eq. (27), the dimensionless number Q_L, is defined as the ratio of viscous forces to the pressure gradient. Subscript L means that the macroscopic length is chosen as the reference scale. Pressure, velocity, density, and temperature in the pores are also used as references to normalize equations:

$$Q_L=O\left(\frac{p_p/L}{\eta\dfrac{\nu_p}{L^2}}\right). \tag{31}$$

An explicit estimation of Q_L is obtained, assuming that the velocity in pores is approached by the velocity in the pores of the fictive single porosity medium,

$$\nu_p=O\left(\frac{|\Pi_p|}{\phi_p\eta}\frac{p_p}{L}\right). \tag{32}$$

Using asymptotic estimations for Π_p, Eq. (31) yields

$$Q_L=\begin{cases}O(\varepsilon^{-2}), & \omega<\omega_\nu,\\ O\left(\varepsilon^{-2}\dfrac{l_p^2}{\delta_\nu^2}\right), & \omega>\omega_\nu.\end{cases} \tag{33}$$

The transient Reynolds number is defined as the ratio between inertial and viscous terms,

$$R_{tL}=O\left(\frac{\rho_0\omega\nu_p}{\eta\dfrac{\nu_p}{L^2}}\right)=O\left(\frac{L^2}{\delta_\nu^2}\right). \tag{34}$$

Air in pores and micropores is a compressible gas, governed by the mass balance equation:

$$j\omega\rho_i+\rho_0\boldsymbol{\nabla}\cdot\boldsymbol{\nu}_i=0 \quad (i=m,p). \tag{35}$$

The Strouhal number needs to be estimated:

$$S_{tL}=O\left(\frac{\omega\rho_p}{\rho_0\nu_p/L}\right). \tag{36}$$

Variations of the volume of air in pores occur at the macroscopic scale:

$$|\boldsymbol{\nabla}\cdot\boldsymbol{\nu}_p|=O\left(\frac{\nu_p}{L}\right), \tag{37}$$

so that

$$S_{tL}=O(1). \tag{38}$$

In order to determine the first-order term of the asymptotic development of $\boldsymbol{\nu}_m$, it is necessary to estimate ν_m/ν_p. This is done using the fictive porous and microporous media, for which the local velocities are estimated by [see Eqs. (1) and (17)]:

$$\nu_i=O\left(\frac{|\Pi_i|}{\phi_i\eta}2\pi\frac{p_i}{|\lambda_i|}\right), \quad i=m,p. \tag{39}$$

Because of the pressure continuity on Γ_{sp}, pressure in pores and micropores must be of the same order [$p_m=O(p_p)$]. Then, using Eq. (11), and considering adiabatic exchanges in pores and micropores, the ratio of the estimated local velocities is given by the ratio of wavelengths:

$$\frac{\nu_m}{\nu_p}=O\left(\left|\frac{\lambda_m}{\lambda_p}\right|\right). \tag{40}$$

2. Thermal effects

Considering that air is an ideal gas, the state equation including possible thermal exchanges is

$$\frac{p_i}{P_0}=\frac{\rho_i}{\rho_0}+\frac{\tau_i}{T_0}\quad (i=m,p). \tag{41}$$

To complete the description and take thermal conduction effects into account, the linearized heat equation is used,

$$\kappa\,\Delta\tau_i - j\omega\rho_0 C_p\tau_i = -j\omega p_i\,. \tag{42}$$

For most of the porous materials, it is realistic to admit that the diffusivity ($\kappa_s/\rho_s C_{ps}$) of the solid phase remains very small compared to the one of the air, so that the acoustical temperature of the skeleton is negligible,

$$\tau_{m/\Gamma_{\rm sm}}=0. \tag{43}$$

Furthermore, on the boundary of the microporous domain, the continuity of temperature on $\Gamma_{\rm sp}$ is

$$\tau_{p/\Gamma_{\rm sp}}=\langle\tau_m\rangle_{\Omega_m}=\frac{1}{\Omega_m}\int_{\Omega_{\rm fm}}\tau_m\,{\rm d}\Omega. \tag{44}$$

Dimensionless numbers. Two dimensionless numbers are defined from (42),

$$N_L=O\left(\left|\frac{\omega\rho_0 C_p\tau_p}{\kappa\dfrac{\tau_p}{L^2}}\right|\right). \tag{45}$$

N_L relates inertial and conductivity terms, and can be simply written as

$$N_L=O\left(\frac{L^2}{\delta_t^2}\right). \tag{46}$$

M_{tL} allows us to estimate the ratio between the source and conductivity term,

$$M_{tL}=O\left(\frac{\omega|p_p|}{\kappa\dfrac{|\tau_p|}{L^2}}\right)=O\left(\phi_p\frac{L^2}{|\Theta_p|}\right). \tag{47}$$

Using asymptotic estimations [Eqs. (6) and (7)] for Θ_p yields

$$M_{tL}=\begin{cases} O(\varepsilon^{-2}), & \omega<\omega_\nu,\\ O\left(\varepsilon^{-2}\dfrac{l_p^2}{\delta_t^2}\right), & \omega>\omega_\nu. \end{cases} \tag{48}$$

III. LOW PERMEABILITY CONTRAST

In this section, we focus our attention on the situation of double porosity material presenting a moderate value of ε_0 ($l_m/l_p\approx 10^{-1}$) [see Fig. 3(a)]. The HSP method is applied, considering the frequency range of particular interest, around $\omega_{\nu m}$, where the coupling between pores and micropores should be strong. The flow and thermal conduction problems are successively treated from the microscopic to the macroscopic point of view.

A. Flow equations

Estimating ε from Eq. (23) for $\omega=O(\omega_{\nu m})$ yields

$$\varepsilon=O(\varepsilon_0^2). \tag{49}$$

Estimations of dimensionless numbers are deduced from Eqs. (33) and (34),

$$Q_L=R_{tL}=O(\varepsilon_0^{-6}). \tag{50}$$

Because in this frequency range wavelengths in pores and micropores are of the same order, one also has [Eq. (40)]

$$\nu_m=O(\nu_p). \tag{51}$$

For the sake of simplicity of writing, parameter ε_0 is used in the asymptotic expansions. Then ε is replaced by ε_0^2 according to Eq. (49)

$$g_p=g_p^0+\varepsilon_0^2 g_p^2+\varepsilon_0^4 g_p^4+\cdots,\quad \text{in the pores},$$

$$h_m=h_m^0+\varepsilon_0 h_m^1+\varepsilon_0^2 h_m^2+\cdots,\quad \text{in the micropores}.$$

Dimensionless operators have to be rewritten, taking the three space variables into account:

$$\boldsymbol{\nabla}=\varepsilon_0^{-1}\varepsilon^{-1}\boldsymbol{\nabla}_z+\varepsilon^{-1}\boldsymbol{\nabla}_y+\boldsymbol{\nabla}_x,\quad \text{for the gradient},$$

$$\Delta=\varepsilon_0^{-2}\varepsilon^{-2}\Delta_z+\cdots,\quad \text{for the Laplacian}$$

(other terms are useless for identification),

it yields

$$\boldsymbol{\nabla}=\varepsilon_0^{-3}\boldsymbol{\nabla}_z+\varepsilon_0^{-2}\boldsymbol{\nabla}_y+\boldsymbol{\nabla}_x\ \text{ and }\ \Delta=\varepsilon_0^{-6}\Delta_z+\cdots.$$

In pores and micropores, normalized Navier–Stokes and mass balance equations are written as

$$\varepsilon_0^6\eta\,\Delta\boldsymbol{\nu}_i-\boldsymbol{\nabla}p_i=j\omega\rho_0\boldsymbol{\nu}_i,$$

$$j\omega\rho_i+\rho_0\boldsymbol{\nabla}\cdot\boldsymbol{\nu}_i=0. \tag{52}$$

1. In the micropores

Introducing asymptotic expansions in normalized equations, and identifying the terms of the same order, one gets

$$O(\varepsilon_0^{-3}):\ \boldsymbol{\nabla}_z p_m^0=\mathbf{0}, \tag{53a}$$

$$O(\varepsilon_0^{-2}):\ \boldsymbol{\nabla}_z p_m^1+\boldsymbol{\nabla}_y p_m^0=\mathbf{0}, \tag{53b}$$

$$O(\varepsilon_0^{-1}):\ \boldsymbol{\nabla}_z p_m^2+\boldsymbol{\nabla}_y p_m^1=\mathbf{0}, \tag{53c}$$

$$O(1):\ \eta\,\Delta_z\boldsymbol{\nu}_m^0-\boldsymbol{\nabla}_z p_m^3-\boldsymbol{\nabla}_y p_m^2-\boldsymbol{\nabla}_x p_m^0=j\omega\rho_0\boldsymbol{\nu}_m^0, \tag{53d}$$

from the Navier–Stokes equation, and

$$O(\varepsilon_0^{-3}):\ \boldsymbol{\nabla}_z\cdot\boldsymbol{\nu}_m^0=0, \tag{54a}$$

$$O(\varepsilon_0^{-2}):\ \boldsymbol{\nabla}_z\cdot\boldsymbol{\nu}_m^1+\boldsymbol{\nabla}_y\cdot\boldsymbol{\nu}_m^0=0, \tag{54b}$$

$$O(\varepsilon_0^{-1}):\ \boldsymbol{\nabla}_z\cdot\boldsymbol{\nu}_m^2+\boldsymbol{\nabla}_y\cdot\boldsymbol{\nu}_m^1=0, \tag{54c}$$

$$O(1):\ j\omega\rho_m^0+\rho_0\boldsymbol{\nabla}_z\cdot\boldsymbol{\nu}_m^3+\rho_0\boldsymbol{\nabla}_y\cdot\boldsymbol{\nu}_m^2+\rho_0\boldsymbol{\nabla}_x\cdot\boldsymbol{\nu}_m^0=0, \tag{54d}$$

from the mass balance equation.

Using Eqs. (53d)–(53c), and the Ω_m periodicity of p_m yields

$$p_m^0=p_m^0(\mathbf{x}), \tag{55a}$$

$$p_m^1=p_m^1(\mathbf{x}), \tag{55b}$$

$$p_m^2=p_m^2(\mathbf{x},\mathbf{y}). \tag{55c}$$

These first results show that the pressure (first order) in micropores varies only according to the macroscopic scale.

The problem consisting of Eq. (53d) and Eq. (54a), including the boundary condition ($\boldsymbol{\nu}^0_{m/\Gamma_{sm}}=\mathbf{0}$), is very similar to the dynamic permeability problem solved by Auriault *et al.*[16] (Sec. II A), in the case of a single porosity medium. Hence, using the equivalent variational formulation, this is a well-posed problem with the existence and uniqueness of the solution, giving the velocity in micropores in a first approximation,

$$\boldsymbol{\nu}^0_m=-\frac{[\pi_m(\mathbf{z},\omega)]}{\eta}(\boldsymbol{\nabla}_y p^2_m+\boldsymbol{\nabla}_x p^0_m). \tag{56}$$

The average velocity over the REVm is introduced, giving the mesoscopic flow law in the microsporous domain,

$$\langle\boldsymbol{\nu}^0_m\rangle_{\Omega_m}=-\frac{[\Pi_m(\omega)]}{\eta}(\boldsymbol{\nabla}_y p^2_m+\boldsymbol{\nabla}_x p^0_m). \tag{57}$$

Note that the permeability tensor $[\Pi_m]$, is the one of the ''fictive'' microsporous media, and it is independent from the mesoscopic structure.

Using the Ω_m periodicity of $\boldsymbol{\nu}^1_m$, the integration of (54b) over the REV leads to the incompressibility relation at the mesoscale,

$$\boldsymbol{\nabla}_y\cdot\langle\boldsymbol{\nu}^0_m\rangle_{\Omega_m}=0. \tag{58}$$

Integrating (54d) over Ω_m yields the mesoscopic balance equation,

$$j\omega\langle\rho^0_m\rangle_{\Omega_m}+\rho_0\boldsymbol{\nabla}_y\cdot\langle\boldsymbol{\nu}^2_m\rangle_{\Omega_m}+\rho_0\boldsymbol{\nabla}_x\cdot\langle\boldsymbol{\nu}^0_m\rangle_{\Omega_m}=0. \tag{59}$$

The behavior of the microsporous medium is now completed from the mesoscopic point of view.

2. In the pores

The identification process yields

$$\boldsymbol{\nabla}_y p^0_p=\mathbf{0}, \tag{60a}$$

$$-\boldsymbol{\nabla}_y p^2_p-\boldsymbol{\nabla}_x p^0_p=j\omega\rho_0\boldsymbol{\nu}^0_p, \tag{60b}$$

$$\boldsymbol{\nabla}_y\cdot\boldsymbol{\nu}^0_p=0, \tag{60c}$$

$$j\omega\rho^0_p+\rho_0\boldsymbol{\nabla}_y\cdot\boldsymbol{\nu}^2_p+\rho_0\boldsymbol{\nabla}_x\cdot\boldsymbol{\nu}^0_p=0, \tag{60d}$$

with the following boundary conditions:

$$\boldsymbol{\nu}^0_{p/\Gamma_{sp}}=\langle\boldsymbol{\nu}^0_m\rangle_{\Omega_m}, \tag{61a}$$

$$\boldsymbol{\nu}^2_{p/\Gamma_{sp}}=\langle\boldsymbol{\nu}^2_m\rangle_{\Omega_m}. \tag{61b}$$

Note that in the frequency range considered ($\omega\gg\omega_{vp}$), the viscous term does not appear in Eq. (60b).

Equation (60a) gives $p^0_p=p^0_p(\mathbf{x})$, leading to show that, from the mesoscopic point of view, the pressure is uniform in the pores at the mesoscopic scale:

$$p^0_m=p^0_p=p^0_{db}(\mathbf{x}). \tag{62}$$

The pressure continuity condition also gives

$$p^2_m(\mathbf{x},\mathbf{y})=p^2_p(\mathbf{x},\mathbf{y}),\quad \text{on}\ \ \Gamma_{sp}. \tag{63}$$

Combining mesoscopic equations [(57), (58)] in the microsporous domain, and in the pores (60) the problem to be solved is rewritten as

$$-\boldsymbol{\nabla}_y p^2_m-\boldsymbol{\nabla}_x p^0_m=\eta[\Pi_m]^{-1}\boldsymbol{\nu}^0_m,\quad \text{on}\ \ \Omega_{sp}, \tag{64a}$$

$$-\boldsymbol{\nabla}_y p^2_p-\boldsymbol{\nabla}_x p^0_p=j\omega\rho_0\boldsymbol{\nu}^0_p,\quad \text{on}\ \ \Omega_{fp}, \tag{64b}$$

$$\boldsymbol{\nabla}_y.\langle\boldsymbol{\nu}^0_m\rangle_{\Omega_m}=0,\quad \text{on}\ \ \Omega_{sp}, \tag{64c}$$

$$\boldsymbol{\nabla}_y.\boldsymbol{\nu}^0_p=0,\quad \text{on}\ \ \Omega_{fp}. \tag{64d}$$

System (64) subjected to boundary conditions (61) is solved using the equivalent variational formulation, and leads to the mesoscopic flow law

$$\boldsymbol{\nu}^0_{db}=-\frac{[\pi_{db}(\mathbf{y},\omega)]}{\eta}\boldsymbol{\nabla}p^0_{db}. \tag{65}$$

The full detailed developments are given in Appendix A.

Finally, the mass balance equation, at the macroscopic scale, is obtained, applying $\langle.\rangle_{\Omega_{sp}}=(1/\Omega_p)\int_{\Omega_{sp}}.\,d\Omega$ to Eq. (59) and $\langle.\rangle_{\Omega_{fp}}=(1/\Omega_p)\int_{\Omega_{fp}}.\,d\Omega$ to Eq. (60d), adding the results and noting that the continuity of flow on Γ_{sp} (using the divergence theorem) yields

$$\int_{\Omega_{sp}}\boldsymbol{\nabla}_y.\langle\boldsymbol{\nu}^2_m\rangle_{\Omega_m}\,d\Omega=-\int_{\Omega_{fp}}\boldsymbol{\nabla}_y\cdot\boldsymbol{\nu}^2_p\,d\Omega. \tag{66}$$

One obtains

$$\frac{j\omega}{\rho_0}(\langle\langle\rho^0_m\rangle_{\Omega_m}\rangle_{\Omega_{sp}}+\langle\rho^0_p\rangle_{\Omega_{fp}})+\boldsymbol{\nabla}_x\cdot\langle\boldsymbol{\nu}^0_{db}\rangle_{\Omega_p}=0. \tag{67}$$

In order to complete the macroscopic description obtained in Eq. (67), and study the macroscopic compressibility of the medium, thermal effects are now included.

B. Compressibility with thermal effects

Estimations of dimensionless numbers are guided by the previous choice of frequency range. Hence, because $\omega_{tm}=O(\omega_{vm})$, Eq. (49) still stands, and one has

$$M_{tL}=N_L=O(\varepsilon_0^{-6}).$$

The normalized equations are

$$\varepsilon_0^6\kappa\,\Delta_z\tau_i-j\omega\rho_0 C_p\tau_i=-j\omega p_i$$

$$\frac{p_i}{P_0}=\frac{\rho_i}{\rho_0}+\frac{\tau_i}{T_0}\quad (i=m,p). \tag{68}$$

1. In the micropores

Introducing asymptotic expansions of microsporous unknowns in Eq. (68) and identifying terms of the same order yields

$$\kappa\,\Delta_z\tau^0_m-j\omega\rho_0 C_p\tau^0_m=-j\omega p^0_m, \tag{69a}$$

$$\frac{p^0_m}{P_0}=\frac{\rho^0_m}{\rho_0}+\frac{\tau^0_m}{T_0}, \tag{69b}$$

$$\tau^0_m=0\ \ \text{on}\ \ \Gamma_{sm}. \tag{69c}$$

Remembering that pressure varies at the macroscopic scale, the solution of Eq. (69a) with (69c) in is given by[17]

$$\tau_m^0 = \frac{\theta_m(\mathbf{z},\omega)}{\kappa} j\omega p_{\mathrm{db}}^0. \tag{70}$$

The mesoscopic temperature law in the microsporous REV is then given by

$$\langle \tau_m^0 \rangle_{\Omega_m} = \frac{\Theta_m(\omega)}{\kappa} j\omega p_{\mathrm{db}}^0, \tag{71}$$

where Θ_m is the thermal permeability of the microsporous material. The average density is obtained from Eq. (69b),

$$\langle \rho_m^0 \rangle_{\Omega_m} = \rho_0 \frac{p_{\mathrm{db}}^0}{K_m(\omega)}, \tag{72}$$

and finally,

$$\langle \langle \rho_m^0 \rangle_{\Omega_m} \rangle_{\Omega_{\mathrm{sp}}} = (1-\phi_p)\rho_0 \frac{p_{\mathrm{db}}^0}{K_m(\omega)}; \tag{73}$$

K_m is the dynamic bulk modulus of the microsporous medium and is absolutely independent from the mesostructure.

2. In the pores

The identification process yields

$$\begin{aligned} &-j\omega\rho_0 C_p \tau_p^0 = -j\omega p_p^0, \\ &\frac{p_p^0}{P_0} = \frac{\rho_p^0}{\rho_0} + \frac{\tau_p^0}{T_0}. \end{aligned} \tag{74}$$

Using $(\rho_0(C_p - C_v) = P_0/T_0)$, the adiabatic relation between density and pressure in pores reads as

$$\rho_p^0 = \rho_0 \frac{p_p^0(\mathbf{x})}{\gamma P_0} \tag{75}$$

and

$$\langle \rho_p^0 \rangle_{\Omega_{\mathrm{fp}}} = \rho_0 \frac{p_{\mathrm{db}}^0}{K_p}, \quad \text{with } K_p = \gamma P_0/\phi_p. \tag{76}$$

Because of the separation of scales, thermodynamic exchanges are adiabatic in this frequency range. However, the study of the low-frequency behavior $[\omega = O(\omega_p)]$ leads to the same relation with $K_p(\omega)$, including conduction effects in pores.[15] Note that $K_p(\omega)$ is the dynamic bulk modulus of the "fictive" porous medium (without a microsporous skeleton), and only depends on the mesostructure.

C. Macroscopic wave equation

Including compressibility with thermal conduction effects [Eqs. (73) and (76)], the macroscopic wave equation can finally be written as

$$j\omega \frac{p_{\mathrm{db}}^0}{K_{\mathrm{db}}} - \boldsymbol{\nabla}_x \cdot \left(\frac{[\Pi_{\mathrm{db}}(\omega)]}{\eta} \boldsymbol{\nabla}_x p_{\mathrm{db}}^0 \right) = 0, \tag{77}$$

with the macroscopic flow law

$$\langle \boldsymbol{\nu}_{\mathrm{db}}^0 \rangle_{\Omega_p} = -\frac{[\Pi_{\mathrm{db}}(\omega)]}{\eta} \boldsymbol{\nabla}_x p_{\mathrm{db}}^0, \tag{78}$$

where $[\Pi_{\mathrm{db}}(\omega)] = (1/\Omega_p)\int_{\Omega_p} [\pi_{\mathrm{db}}(\mathbf{y},\omega)]\,\mathrm{d}\Omega$.

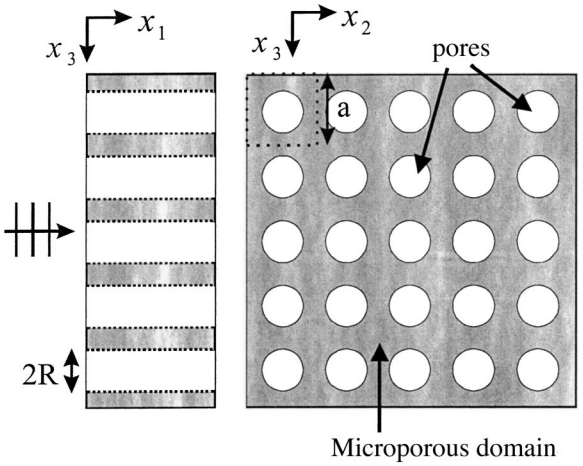


FIG. 4. Periodic perforated microporous material.

This law looks like the generalized Darcy's law, found in single porosity media, however, properties of the dynamic permeability tensor $[\Pi_{\mathrm{db}}(\omega)]$ are slightly different: flow in the pores is governed by inertial forces, whereas it is viscoinertial in the micropores.

This macroscopic description is valid until ω is large compared to $\omega_{\nu p}$. The study of the low-frequency behavior $[\omega = O(\omega_{\nu p})]$ also leads to Eq. (67). However, in this frequency range, the wavelength in pores is large compared to the one in micropores, so that the macroscopic flow is given by the flow in the pores ($|\nu_p| \gg |\nu_m|$ and $|\Pi_p| \gg |\Pi_m|$). The reader can refer to the work of Royer[13] for more details. Then, the flow law is

$$\langle \boldsymbol{\nu}_{\mathrm{db}}^0 \rangle_{\Omega_p} = \langle \boldsymbol{\nu}_p^0 \rangle_{\Omega_p} = -\frac{[\Pi_p(\omega)]}{\eta} \boldsymbol{\nabla}_x p_{\mathrm{db}}^0. \tag{79}$$

Hence, Eq. (78) can be extended to the whole frequency range, noting that $[\Pi_{\mathrm{db}}(\omega)]$ tends to $[\Pi_p(\omega)]$ as ω tends to 0.

Determining $[\Pi_{\mathrm{db}}(\omega)]$ involves that Eqs. (64) have to be solved for a given mesogeometry. In Sec. III D, we present an example of simple geometry for which an analytical expression can be found.

In Eq. (77) the macroscopic bulk modulus of the double porosity medium appears to be given by

$$K_{\mathrm{db}}(\omega) = \left(\frac{1}{K_p(\omega)} + (1-\phi_p)\frac{1}{K_m(\omega)} \right)^{-1}. \tag{80}$$

Contrary to the flow problem, it is here possible to give the exact expression for K_{db} with respect to the dynamic bulk modulus of the "fictive" porous and microporous media.

D. Example: perforated microporous material

In order to illustrate theoretical results obtained in Secs. III A and III B, consider the case of a plane wave propagating in a perforated microporous material (Fig. 4). The circular cylindrical holes are parallel to the direction of propagation $(\mathbf{x}_1)$, and constitutes the pores network. These holes are placed in a regular square lattice, so that the mesoscopic geometry shows a simple squared periodicity in the $(\mathbf{x}_2, \mathbf{x}_3)$

plane. The period is arbitrary in the $\mathbf{x}_1$ direction, as the material is supposed to be of infinite length. According to the meso-geometry, ϕ_p is $\pi R^2/a^2$. The purpose is to calculate analytically the permeability and the bulk modulus of the medium.

1. Permeability

At the mesoscale, Eqs. (64a) and (64b) in the direction of $\mathbf{x}_1$ become

$$\langle \boldsymbol{\nu}_m^0 \rangle_{\Omega_m} . \mathbf{x}_1 = -\frac{\Pi_m^{x_1}}{\eta}\left(\frac{\partial p_m^2}{\partial y_1} + \frac{\partial p_{\mathrm{db}}^0}{\partial x_1}\right),$$
$$\boldsymbol{\nu}_p^0 . \mathbf{x}_1 = -\frac{1}{j\omega\rho_0}\left(\frac{\partial p_p^2}{\partial y_1} + \frac{\partial p_{\mathrm{db}}^0}{\partial x_1}\right). \tag{81}$$

$\Pi_m^{x_1}$ is the dynamic permeability of the microporous medium along $\mathbf{x}_1$. Noting that $(\partial./\partial y_1 = 0)$, the macroscopic flow is

$$\langle \boldsymbol{\nu}_{\mathrm{db}}^0 \rangle . \mathbf{x} = -\frac{\Pi_{\mathrm{db}}^{x_1}(\omega)}{\eta}\frac{\partial p_{\mathrm{db}}^0}{\partial x_1}, \tag{82}$$

with

$$\Pi_{\mathrm{db}}^{x_1}(\omega) = \frac{\eta\phi_p}{j\omega\rho_0} + (1-\phi_p)\Pi_m^{x_1}(\omega). \tag{83}$$

Note that this expression is a simple parallel models for flow, and can be extended to low frequencies adding viscoinertial effects in pores. Hence, a wide-frequency range expression, checking both low- and middle-frequency solutions (in a first approximation) is proposed:

$$\Pi_{\mathrm{db}}^{x_1}(\omega) = (1-\phi_p)\Pi_m^{x_1}(\omega) + \Pi_p^{x_1}(\omega). \tag{84}$$

It is common in acoustics to use the complex dynamic equivalent density of the homogeneous medium, defined as

$$\rho_{\mathrm{db}}^{x_1} = \frac{\eta}{j\omega\Pi_{\mathrm{db}}^{x_1}} = ((1-\phi_p)/\rho_m^{x_1} + 1/\rho_p^{x_1})^{-1}. \tag{85}$$

It is to be noticed that Eqs. (84) and (85) remain valid for all mesoscopic geometry with straight parallel cylindrical or slit pores.

Using the Zwikker and Kosten[6] expression for this particular geometry, the permeability of the porous material, in the $\mathbf{x}_1$ direction is given by

$$\Pi_p^{x_1}(\omega) = \frac{\phi_p}{j\delta_\nu^2}\left(1 - \frac{2}{\mu\sqrt{-j}}\frac{J_1(\mu\sqrt{-j})}{J_0(\mu\sqrt{-j})}\right), \tag{86}$$

where $\mu = R/\delta_\nu$ is a dimensionless parameter and J_0 and J_1 are the Bessel functions of order 0 and 1. Note that for the porous material, the static permeability is $\Pi_p(0) = \phi_p R^2/8$, and the viscous characteristic frequency is $\omega_{\nu p} = 8\eta/\rho_0 R^2$.

The dynamic permeability of the single porosity medium can be computed using, for instance, the semiphenomenological model of Johnson *et al.*,[19] which gives

$$\Pi(\omega) = \Pi(0) \Big/ \left(\frac{j\omega}{\omega_\nu} + \left(1 + j\frac{M}{2}\omega_\nu\right)^{1/2}\right), \tag{87}$$

TABLE I. Parameters used to compute the dynamic permeabilities of the porous and the microporous media constituting the double porosity medium.

	$\Pi(0)$ (m^2)	ω_ν (rad s^{-1})	M
Porous medium	2.5×10^{-7}	31 (=5 Hz)	1
Microporous medium	3.1×10^{-9}	4800 (=764 Hz)	1

with $\omega_\nu = \eta\phi/\rho_0\alpha_\infty\Pi(0)$ and $M = 8\alpha_\infty\Pi(0)/\phi\Lambda^2$, with Λ the viscous characteristic length.

In the following example, parameters have been chosen to approach the identified "low contrast of permeability" situation (see Table I), with $R = 0.002$ m and $a = 0.005$ m.

Figure 5 illustrates the evolution of the permeability of the double porosity medium with frequency. At low frequency, the microporous domain does not strongly modify the global flow. It is especially important to note that Π_{db} is increased in the middle-frequency range, as the flow is simultaneously inertial in the pores and viscoinertial in the micropores. The existence and the influence of the two characteristic frequencies appear clearly on these graphs.

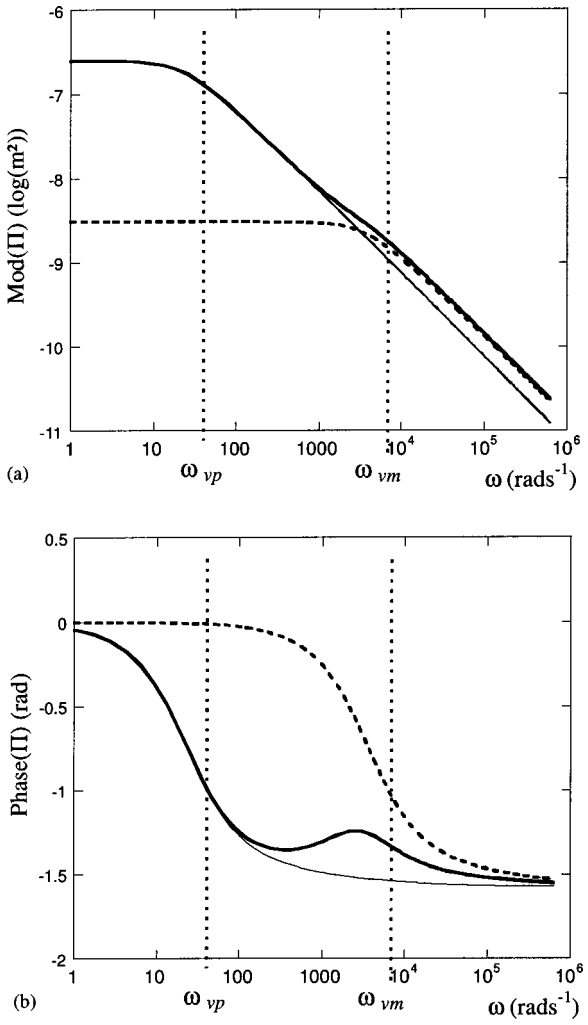


FIG. 5. Low permeability contrast situation. Modulus (a) and phase (b) of, Π_p (thin line), Π_m (dashed line), Π_{db} (thick line).

TABLE II. Parameters used to compute the bulk modulus of the porous and the microporous media constituting the double porosity medium.

	$\Theta(0)$ (m^2)	ω_t (rad s^{-1})	M
Porous medium	2.5×10^{-7}	43 (=7 Hz)	1
Microporous medium	4.3×10^{-9}	4800 (=764 Hz)	1

2. Bulk modulus

Considering the circular cylindrical geometry, the thermal permeability function related to the network is

$$\Theta_p(\omega)=\frac{\phi_p}{j\delta_t^2}\left(1-\frac{2}{\mu\sqrt{-j}}\frac{J_1(\mu\sqrt{-j})}{J_0(\mu\sqrt{-j})}\right); \tag{88}$$

K_p is computed, introducing Eq. (88) in (9) (Ref. 3, p. 90).

A semiphenomelogical model is used to compute Θ_m and K_m,

$$\Theta(\omega)=\Theta(0)\Bigg/\left(\frac{j\omega}{\omega_t}+\left(1+j\frac{M'}{2}\omega_t\right)^{1/2}\right), \tag{89}$$

with $\omega_t=\kappa\phi/\rho_0 C_p\Theta(0)$ and $M'=8\Theta(0)/\phi\Lambda'^2$, where Λ' is the thermal characteristic length.[3,17]

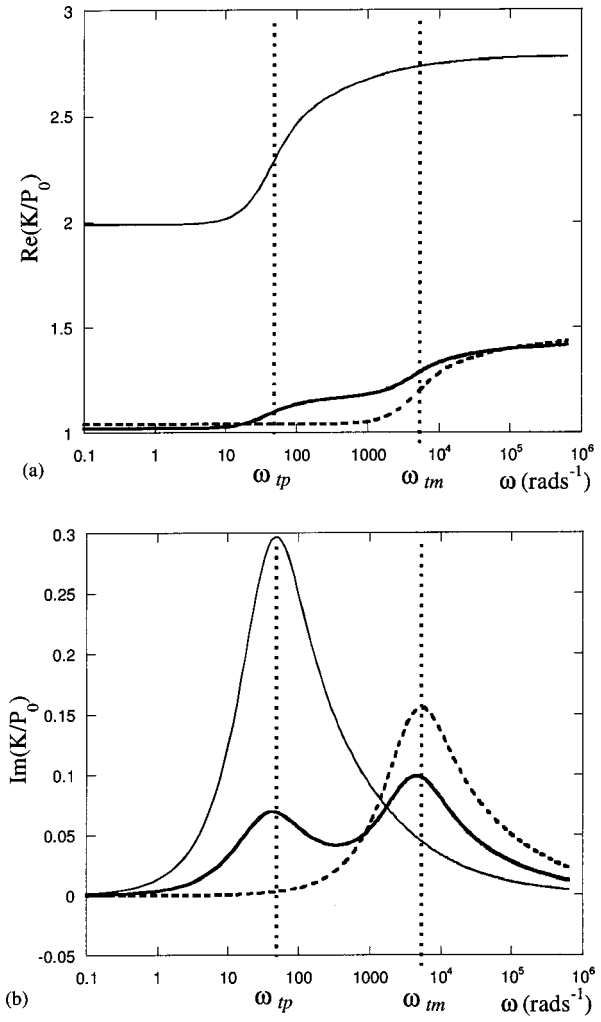


FIG. 6. Low permeability contrast situation. Real (a) and imaginary (b) parts of K_p/P_0 (thin line), K_m/P_0 (dashed line), K_{db}/P_0 (thick line).

Parameters used to evaluate the K_i are given in Table II.

Regarding compressibility and thermal effects, Fig. 6 shows the ''double characteristic frequencies behavior'' involved by the separation of thermal characteristic frequencies between pores and micropores. Thermal exchanges in the equivalent homogeneous medium are successively isothermal at low frequency ($|K_{db}|\approx P_0/\phi_{db}$, for $\omega\ll\omega_{tp}$), they become adiabatic in the pores, in the middle-frequency range ($|K_{db}|\approx\gamma P_0/(\phi_p+\gamma(1-\phi_p)\phi_m)$, for $\omega_{tp}<\omega<\omega_{tm}$), and are finally adiabatic in the entire system at higher frequencies ($|K_{db}|\approx\gamma P_0/\phi_{db}$, for $\omega\gg\omega_{tm}$). The imaginary part of K_{db} shows two local maximums, corresponding approximately to ω_{tp}, and ω_{tm}.

IV. HIGH PERMEABILITY CONTRAST

This second situation, identified in Fig. (3b), occurs for a small value of the interscale ratio ($\varepsilon_0\approx10^{-3}$). The key point of this case is linked to the fact that the wavelength in the microporous domain is of the same order of magnitude as the mesoheterogeneities. This phenomenon was presupposed and treated by Boutin *et al.*,[14] but needed to be justified in the context of ''realistic'' materials. In this section, the main results are exposed and discussed, readers will refer to Appendix B for more details.

A. Macroscopic flow

In the whole frequency range of homogenization, the wavelength in the pores is large compared to the one in the micropores, so that the flow in the microporous medium does not modify the macroscopic flow:

$$\langle\boldsymbol{\nu}_{db}^0\rangle_{\Omega_p}=\langle\boldsymbol{\nu}_p^0\rangle_{\Omega_p}=-\frac{[\Pi_p(\omega)]}{\eta}\nabla_x p_{db}^0(\mathbf{x}), \tag{90}$$

and the macroscopic pressure is given by the pressure in the pores $[p_{db}^0=p_p^0(\mathbf{x})]$.

B. Pressure diffusion effects

Considering the frequency range around ω_d (see Sec. II C), it appears that the pressure in the micropores does not only vary at the macroscale, but also at the mesoscale. Moreover, because ω_d is small compared to $\omega_{\nu m}$, the flow in the micropores is mainly viscous $[\Pi_m(\omega)\approx\Pi_m(0)]$. The wave equation governing pressure in the REV (Ω_m) is simply given by the following diffusion equation [Eq. (B9)]:

$$j\omega\phi_m\frac{p_m^0(\mathbf{x},\mathbf{y})}{P_0}-\frac{\Pi_m(0)}{\eta}\Delta_y p_m^0(\mathbf{x},\mathbf{y})=0,\quad\text{in }\Omega_{sp}. \tag{91}$$

The key point of this situation is that the boundary condition is imposed by the pressure in the pores $[p_m(\mathbf{x},\mathbf{y})=p_p(\mathbf{x})$ on $\Gamma_{sp}]$ that varies at macroscopic scale. This is the direct consequence of the fact that the wavelength in the pores is greater than in the micropores. The problem to be solved is then very similar to the one of thermal diffusion effects in the fluid network of a porous material with an isothermal skeleton. Hence, p_m^0 is linearly related to p_p^0,

$$p_m^0=\left(1-j\omega\frac{\phi_m\eta}{P_0\Pi_m(0)}d(\omega,\mathbf{y})\right)p_p^0. \tag{92}$$

Herein, d is a scalar function depending only on the mesogeometry and the frequency, and it has the dimension of a permeability (m^2). From Eq. (91), the length δ_d is introduced, and defined as the pressure diffusion skin depth:

$$\delta_d = \sqrt{\frac{P_0 \Pi_m(0)}{\phi_m \eta \omega}}; \tag{93}$$

δ_d is analog to the viscous or thermal skin depths, commonly defined from the Navier–Stokes and the conduction equation. In the pressure diffusion problem, δ_d gives, at high frequency, an estimation of the boundary layer in which p_m^0 varies strongly.

Let the function F_d be the ratio of the average pressure in the microporous domain $[\langle\cdot\rangle_{\tilde{\Omega}_{sp}} = (1/\Omega_{sp})\int_{\Omega_{sp}} \cdot \, d\Omega]$, to the pressure in pores:

$$F_d(\omega) = \frac{\langle p_m^0 \rangle_{\tilde{\Omega}_{sp}}}{p_p^0} = \left(1 - j\omega \frac{\phi_m \eta}{P_0 \Pi_m(0)(1-\phi_p)} D(\omega)\right), \tag{94}$$

which can be rewritten as

$$F_d(\omega) = \left(1 - j \frac{\omega}{\omega_d} \frac{D(\omega)}{D(0)}\right), \tag{95}$$

with D defined as $D(\omega) = (1/\Omega_p)\int_{\Omega_{sp}} d(\mathbf{y}, \omega) d\Omega$.

Using again the analogy with the thermal conduction problem, properties of D can be directly compared to the ones of the "thermal permeability" Θ [see Eq. (5)] introduced by Lafarge *et al.*[17] As it is shown in Sec. IV D, Eq. (91) can be solved for simple mesoscopic geometries, and analytical expressions for D can be derived. For more complex geometries, a semiphenomenological model can, however, be used. The model proposed here is based on the one given by Johnson *et al.* for the permeability, and also used by Lafarge to describe thermal effects:

$$D(\omega) = D(0) \Big/ \left(\frac{j\omega}{\omega_d} + \left(1 + j \frac{M_d}{2} \omega/\omega_d\right)^{1/2}\right). \tag{96}$$

$D(0)$, the static value for D, is a geometrical parameter, comparable to the "static thermal permeability" of Lafarge *et al.* It can be estimated with $D(0) = O((1-\phi_p) l_p^2)$.

ω_d is defined as the characteristic frequency of pressure diffusion effects, and can be expressed according to the properties of the microporous material and the mesogeometry:

$$\omega_d = \frac{(1-\phi_p) P_0 \Pi_m(0)}{\phi_m \eta D(0)}. \tag{97}$$

This last expression should be related to the estimation given in Eq. (25). ω_d separates two frequency ranges.

(i) At low frequency ($\omega \ll \omega_d$): F_d tends to 1 as frequency decreases. At the mesoscopic scale pressure is uniform in the whole material.

(ii) At high frequency ($\omega \gg \omega_d$): F_d, and then $\langle p_m^0 \rangle_{\tilde{\Omega}_{sp}}$, tends to 0 as frequency increases, meaning that the micropores do not participate in the macroscopic behavior of the material anymore.

The most interesting behavior is, of course, obtained around ω_d when the pressure in pores and micropores are of the same order. Because F is complex, it is to be underlined that $\langle p_m^0 \rangle_{\tilde{\Omega}_{sp}}$ and p_p^0 present a phase mismatch, inducing a new dissipation effect that does not exist in simple porosity media.

The form parameter M_d introduced in the model can be related to Λ_d, defined as the characteristic length of the microporous domain:

$$M_d = \frac{8}{\Lambda_d^2} \frac{D(0)}{(1-\phi_p)}, \tag{98}$$

with

$$\Lambda_d = 2 \frac{\Omega_{sp}}{\Gamma_{sp}}. \tag{99}$$

One will notice that Λ_d is directly related to the thermal characteristic length of the pores network ($\Lambda_p' = 2\Omega_{fp}/\Gamma_{sp}$)[3] as follows:

$$\Lambda_d = \frac{(1-\phi_p)}{\phi_p} \Lambda_p'. \tag{100}$$

C. Macroscopic wave equation

The wave equation in the equivalent homogeneous medium, assuming that air in pores and micropores remains isothermal, is given by (B18):

$$j\omega \frac{p_p^0}{K_{db}} - \boldsymbol{\nabla}_x \cdot \left(\left[\frac{\Pi_p(\omega)}{\eta} \boldsymbol{\nabla} p_p^0\right]\right) = 0, \tag{101}$$

with

$$K_{db}(\omega) = \left(\frac{\phi_p}{P_0} + (1-\phi_p) \frac{\phi_m F(\omega)}{P_0}\right)^{-1}. \tag{102}$$

Adding thermal effects, a more complete expression was proposed by Boutin *et al.*:[14]

$$K_{db}(\omega) = \left[\frac{1}{K_p} + (1-\phi_p) \frac{F_d\left(\omega \dfrac{P_0}{\phi_m K_m}\right)}{K_m}\right]^{-1}. \tag{103}$$

Comparing Eqs. (103) and (80) shows that the consequence of the pressure diffusion effects is a partial coupling ($|F_d| \leqslant 1$) between pores and micropores, which tends to increase the apparent bulk modulus as frequency increases.

In the following paragraph, theoretical examples are presented to illustrate this effect, in the case of simple mesoscopic pores structures.

D. Examples

In this section, two examples are presented, to show how the pressure diffusion effects are calculated, and to point out the influence of the mesogeometry on the macroscopic behavior of the double porosity material.

1. Slits materials

First, consider a double porosity material made of parallel microporous panels of thickness $2b$ separated with slits of thickness $2a$ (Fig. 7), and a meso-porosity $\phi_p = a/(a+b)$.

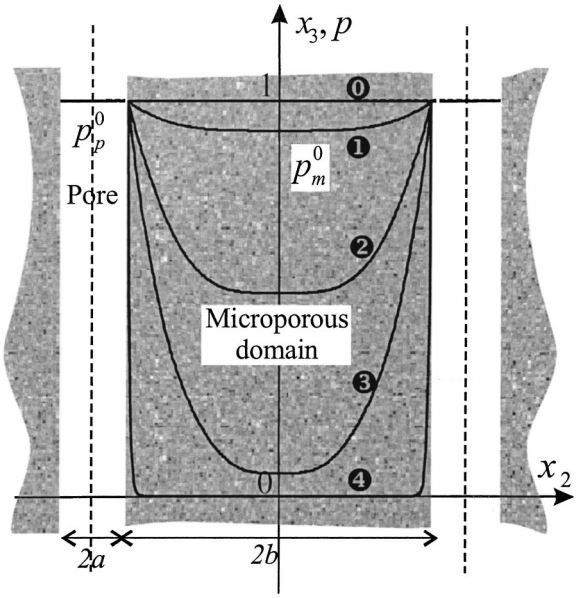


FIG. 7. High permeability contrast situation. Evolution of the pressure field in the microporous domain when frequency increases: ⓪ $q=0.1$, ❶ $q=1$, ❷ $q=2$, ❸ $q=5$, ❹ $q=100$, ($q=b/\delta_d$).

Because of the simplicity of the geometry, the mesoscopic function d, introduced in Eq. (92), can be analytically calculated (Ref. 3, p. 51):

$$d(x_2)=-j\delta_d^2\left(1-\frac{\cosh\left(\frac{x_2}{\delta_d}\sqrt{j}\right)}{\cosh\left(\frac{b}{\delta_d}\sqrt{j}\right)}\right). \tag{104}$$

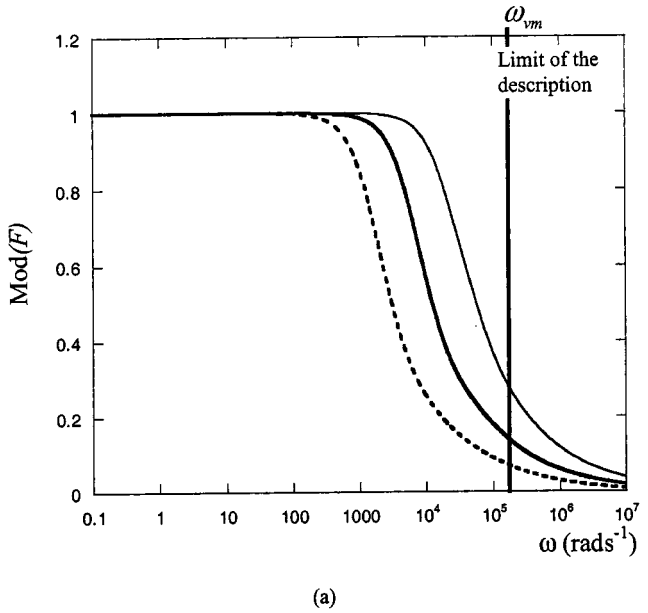
In Fig. 7 a representation of the mesoscopic pressure field in the microporous domain for different value of $q=b/\delta_d$ is depicted. For sufficiently low pulsations (small values of q), the pressure is almost uniform in $\Omega_{\rm sp}$, whereas it is much smaller than p_p^0 as the solicitations becomes faster, except in the boundary layer.

From the macroscopic point of view, function F is calculated with (95), where $D(\omega)$ is obtained integrating Eq. (104) between $-b$ and b:

$$D(\omega)=-j(1-\phi_p)\delta_d^2\left(1-\frac{\tanh\left(\frac{b}{\delta_d}\sqrt{j}\right)}{\frac{b}{\delta_d}\sqrt{j}}\right). \tag{105}$$

TABLE III. Parameters of the diffusion function for three configurations of slit material.

Config.	a (m)	b (m)	$D(0)$ (m^2)	ω_d (rad s^{-1})
1	0.005	0.01	2.2×10^{-5}	30705 (=4887 Hz)
2	0.01	0.02	8.9×10^{-5}	7676 (=1221 Hz)
3	0.02	0.04	3.6×10^{-4}	1920 (=305 Hz)



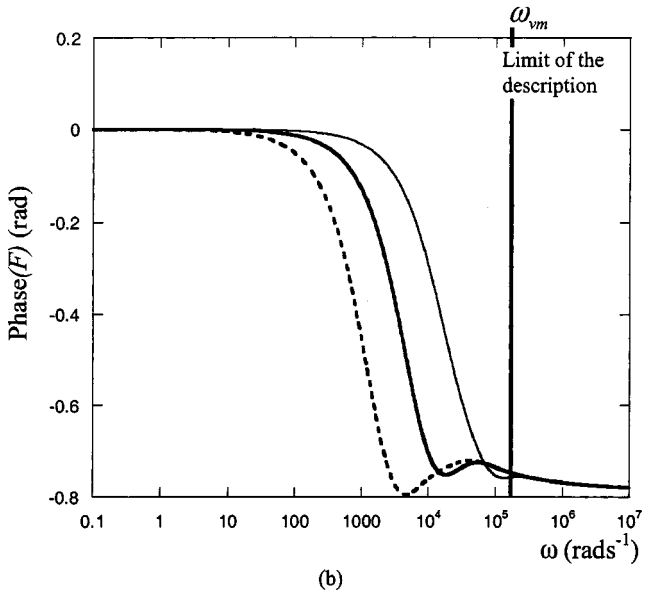


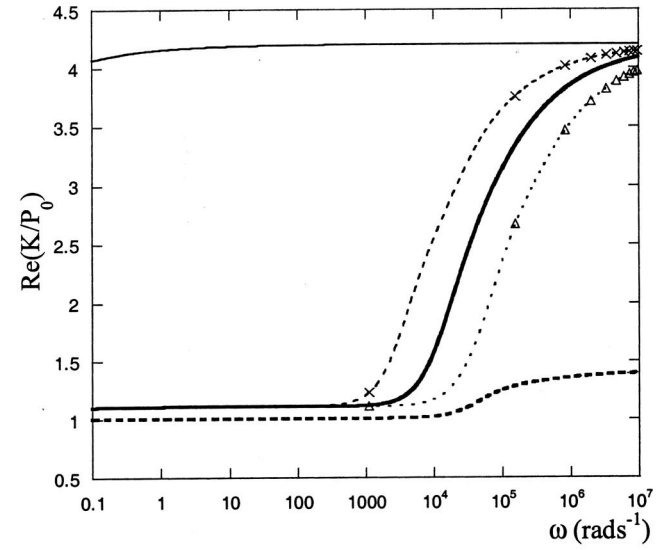
FIG. 8. High permeability contrast situation. Evolution of function F [(a) modulus, (b) phase] with frequency for three mesogeometry: config. 1 (thin line), config. 2 (thick line), config. 3 (dashed line). Refer to Table III for the configurations.

Note that the static value of $D(\omega)$ is $D(0)=(1-\phi_p)\times(b^2/3)$.

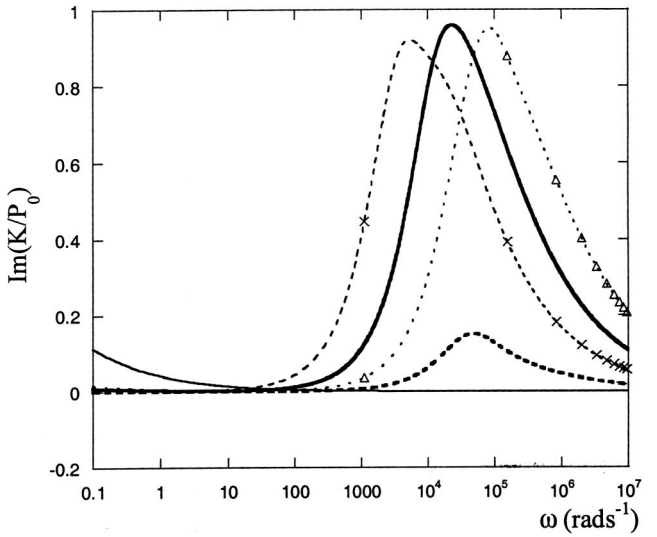
Three configurations are now presented to illustrate the pressure diffusion effects. They are corresponding to three parallel slit materials having the same mesoporosity ($\phi_p=0.33$), but with different values for a and b (Table III).

In Figs. 8 and 9, functions F_d and $K_{\rm db}$ are, respectively, depicted, calculated from Eqs. (95) and (103). K_p and K_m are, respectively, computed using the slits model,[3] and the Lafarge *et al.*-model [Eqs. (9) and (89)]. Parameters of the microporous medium (Table IV) have been chosen in order to have a large value for ω_{vm}, compared to ω_d.

These simulations, that should be compared to the ones obtained in the case of low permeability contrast (Fig. 6), highlights the role played by diffusion effects when the fre-



(a)



(b)

FIG. 9. High permeability contrast situation. Real (a) and imaginary (b) parts of K_p/P_0 (thin line), K_m/P_0 (dashed line), K_{db}/P_0 [config. 1 (dashed line with triangles), config. 2 (thick line), config. 3 (dashed line with crosses)].

quency approach ω_d: the bulk modulus (real and imaginary parts) of the equivalent medium is dramatically increased, because of the decreasing of the average pressure in Ω_{sp}. These graphs also show that the mesoscopic geometry has a great influence on the macroscopic behavior, shifting in a very sensible way the frequency range in which diffusion effects occur.

Important remark: the high permeability contrast description obtained is valid until waves are diffusive in Ω_{sp} (i.e., $\omega \ll \omega_{\nu m}$). As ω approaches $\omega_{\nu m}$, waves becomes propagative in the microporous medium, and the model does not hold anymore. Hence, the high-frequency part of curves presented on the graphs is purely mathematical and has no physical sense. It can also be supposed that in this frequency range the low permeability contrast behavior can be found again as wavelengths in pores and micropores becomes of the same order. However, in this case one must check that homogenization conditions are still verified, and that no diffraction effects appears.

TABLE IV. Acoustical parameters of the microporous material.

$\Pi_m(0)$ (m^2)	$\omega_{\nu m}$ (rad s^{-1})	M
1.85×10^{-10}	82500 (=13130 Hz)	1
$\Theta_m(0)$ (m^2)	ω_{tm} (rad s^{-1})	M'
1.23×10^{-10}	67000 (=10663 Hz)	1

2. Perforated material

This second example deals with the perforated microporous material depicted in Fig. 4. In this case the mesogeometry is much more complicated and no simple analytical expression can be given for the diffusion functions $D(\omega)$ and $F_d(\omega)$. It is, however, possible to use the semiphenomenological model given in Eq. (96). The static value $D(0)$ can actually be calculated using the expression proposed of Tarnow[20] for the determination of the static permeability of a square lattice of fibers. In this case the calculus of $D(0)$ is the same as the one of static permeability in the direction of fibers [see Eq. (B12)]:

$$D(0)=\frac{a^2}{4\pi}\left(\ln\left(\frac{1}{\phi_p}\right)-\frac{3}{2}+2\phi_p-\frac{\phi_p^2}{2}\right). \tag{106}$$

One will note that M_d can be calculated simply using $\Lambda_d=(a^2-\pi R^2)/\pi R$.

V. CONCLUSION

The acoustic behaviors of porous media with a double porosity have been investigated. The homogenization method, applied to the triple-scales heterogeneous materials, revealed the interscale couplings between the two identified fluid networks. Considering materials with realistic characteristics, explicit estimations, especially of wavelengths in porous and microporous media, has been used, and led to identify two macroscopic behaviors, corresponding to materials with very different interscale ratios ($\varepsilon_0=l_m/l_p$).

The first situation, called ''low permeability contrast,'' is the result of a moderate ratio between the characteristic sizes of the pores and the micropores. The macroscopic description obtained with the help of the asymptotic development method, reveals a strong coupling between the two networks. The macroscopic flow involves both networks, and is given by a kind of generalized Darcy's law, associated with the characteristic frequencies of the pores and micropore networks. A total coupling behavior was also found regarding compressibility effects. The macroscopic bulk modulus directly depends on the ones associated with pores and microporous material, and its frequency behavior is narrowly related to the thermal characteristic frequency of the two networks.

The ''high permeability contrast case'' has been studied second, and we have justified that mesoscopic pressure variations can occur, in the microporous part of the double porosity medium, when l_m is small enough compared to l_p. The macroscopic description obtained in this case showed that the macroscopic flow law is only governed by the flow in the pores. The microporous domain does not participate in the macroscopic flow, but it is partially involved in the compressibility of the equivalent medium. At the mesoscopic scale, the pressure in the microporous domain is given by a diffusion equation. It induces a new dissipation effect, directly related to the mesoscopic structure of the material, and to the properties of the microporous part.

In this work we give new perspectives for the design of passive porous absorbers. Creating a mesoscopic geometry (such as holes or slits) in a microporous matrix, can dramatically modify its acoustical behavior and performances. For instance, it is possible to increase and adjust the absorption coefficient of a given microporous medium, on a wide frequency range, modifying the mesoscopic parameters involved in the diffusion function F.[21,22]

APPENDIX A: FLOW LAW IN THE CASE OF LOW PERMEABILITY CONTRAST

The system of equations to solve is

$$\begin{aligned}
&-\boldsymbol{\nabla}_y p_m^2-\boldsymbol{\nabla}_x p_m^0=\eta[\Pi_m]^{-1}\boldsymbol{\nu}_m^0, \quad \text{on } \Omega_{\text{sp}},\\
&-\boldsymbol{\nabla}_y p_p^2-\boldsymbol{\nabla}_x p_p^0=j\omega\rho_0\boldsymbol{\nu}_p^0, \quad \text{on } \Omega_{\text{fp}},\\
&\boldsymbol{\nabla}_y\cdot\langle\boldsymbol{\nu}_m^0\rangle_{\Omega_m}=0, \quad \text{on } \Omega_{\text{sp}},\\
&\boldsymbol{\nabla}_y\cdot\boldsymbol{\nu}_p^0=0, \quad \text{on } \Omega_{\text{fp}}.
\end{aligned} \tag{A1}$$

We search the solution $\boldsymbol{\nu}_{\text{db}}$, defined on Ω_p by

$$\boldsymbol{\nu}_{\text{db}}^0=\begin{cases}\langle\boldsymbol{\nu}_m^0\rangle_{\Omega_m}, & \text{on } \Omega_{\text{sp}},\\ \boldsymbol{\nu}_p^0, & \text{on } \Omega_{\text{fp}},\end{cases} \tag{A2}$$

with

$$\langle\boldsymbol{\nu}_m^0\rangle_{\Omega_m}=\boldsymbol{\nu}_p^0, \quad \text{on } \Gamma_{\text{sp}}. \tag{A3}$$

Using an appropriate Hilbert space of solutions W:

$$\begin{aligned}W=\{&\mathbf{w}\in\Omega_p,\ \Omega_p \text{ périodic}/\boldsymbol{\nabla}_y\cdot\mathbf{w}=0,\\ &\text{continuous on } \Gamma_{\text{sp}}\},\end{aligned} \tag{A4}$$

with its Hermitian product:

$$(\mathbf{u},\boldsymbol{\nu})=\int_{\Omega_{\text{sp}}}\mathbf{u}.\overline{\boldsymbol{\nu}}\,\mathrm{d}\Omega+\int_{\Omega_{\text{fp}}}\mathbf{u}.\overline{\boldsymbol{\nu}}\,\mathrm{d}\Omega, \tag{A5}$$

$\overline{\boldsymbol{\nu}}$ is the conjugate of $\overline{\boldsymbol{\nu}}$.

The ''Lax–Milgram's lemma'' is then used to guarantee the existence and uniqueness of the solution.

The weak formulation of the problem is

$$\begin{aligned}
&-\int_{\Omega_{\text{sp}}}\boldsymbol{\nabla}_y p_m^2\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega-\int_{\Omega_{\text{fp}}}\boldsymbol{\nabla}_y p_p^2\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega-\int_{\Omega_{\text{sp}}}\boldsymbol{\nabla}_x p_{\text{db}}^0\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega\\
&\quad-\int_{\Omega_{\text{fp}}}\boldsymbol{\nabla}_x p_{\text{db}}^0\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega\\
&=\eta[\Pi_{\text{eqm}}^s]^{-1}\int_{\Omega_{\text{sp}}}\boldsymbol{\nu}_{\text{db}}^0\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega+j\omega\rho_0\int_{\Omega_{\text{fp}}}\boldsymbol{\nu}_{\text{db}}^0\overline{\mathbf{w}}\,\mathbf{d}\Omega.
\end{aligned} \tag{A6}$$

Noting that

$$\int_{\Omega_{\text{sp}}}\boldsymbol{\nabla}_y p_m^2\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega=\int_{\Gamma_{\text{sp}}}p_m^2\cdot\overline{\mathbf{w}}\cdot\mathbf{n}\,\mathrm{d}\Gamma-\int_{\Omega_{\text{sp}}}p_m^2\boldsymbol{\nabla}_y\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega$$

($\mathbf{n}$: normal to Γ_{sp}),

$$\int_{\Omega_{\text{fp}}}\boldsymbol{\nabla}_y p_p^2\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega=-\int_{\Gamma_{sp}}p_p^2\cdot\overline{\mathbf{w}}\cdot\mathbf{n}\,\mathrm{d}\Gamma-\int_{\Omega_{\text{fp}}}p_p^2\boldsymbol{\nabla}_y\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega,$$

and $\boldsymbol{\nabla}_y\cdot\overline{\mathbf{w}}=0$.

Adding pressure continuity, terms with p_m^2, p_p^2 disappear in the variational formulation:

$$\begin{aligned}
-\boldsymbol{\nabla}_x p_{\text{db}}^0\int_{\Omega_p}\overline{\mathbf{w}}\,\mathrm{d}\Omega&=\eta[\Pi_m]^{-1}\int_{\Omega_{\text{sp}}}\boldsymbol{\nu}_{\text{db}}^0\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega\\
&\quad+j\omega\rho_0\int_{\Omega_{\text{fp}}}\boldsymbol{\nu}_{\text{db}}^0\cdot\overline{\mathbf{w}}\,\mathrm{d}\Omega.
\end{aligned} \tag{A7}$$

From this equation, one can affirm that $\boldsymbol{\nu}_{\text{db}}^0$, $\boldsymbol{\nabla}_x p_{\text{db}}^0$ are linearly related, and the flow at the mesoscale can be written as

$$\boldsymbol{\nu}_{\text{db}}^0=-\frac{[\pi_{\text{db}}(\mathbf{y},\omega)]}{\eta}\boldsymbol{\nabla}p_{\text{db}}^0. \tag{A8}$$

APPENDIX B: DETAILED CALCULATIONS FOR THE CASE OF HIGH PERMEABILITY CONTRAST

Here, normalized equations are written considering that $\omega=O(\omega_d)$ [see Fig. 3(b)]. In this frequency range we have

$$\varepsilon_0=\varepsilon^3, \tag{B1}$$

and

$$Q_L=R_{tL}=O(\varepsilon^{-6}) \tag{B2}$$

The normalized equations in pores and micropores are

$$\begin{aligned}
&\varepsilon^6\eta\,\Delta\boldsymbol{\nu}_i-\boldsymbol{\nabla}p_i=j\omega\rho_0\boldsymbol{\nu}_i\\
&j\omega\rho_i+\rho_0\boldsymbol{\nabla}\cdot\boldsymbol{\nu}_i=0
\end{aligned} \quad (i=m,p). \tag{B3}$$

1. In micropores

For $\omega=O(\omega_d)$, we saw [Fig. 3(b)] that $|\lambda_m/2\pi|=O(l_p)$; hence, using Eq. (29) yields $\nu_m=O(\varepsilon\nu_p)$. For the sake of simplicity, asymptotic expansions of unknowns are written using ε:

$$\begin{aligned}
\boldsymbol{\nu}_m&=\varepsilon\boldsymbol{\nu}_m^1+\varepsilon^4\boldsymbol{\nu}_m^4+\cdots,\\
p_m&=p_m^0+\varepsilon^3p_m^3+\cdots,\\
\rho_m&=\rho_m^0+\varepsilon^3\rho_m^3+\cdots.
\end{aligned}$$

The normalized operators are written as

$$\begin{aligned}
\boldsymbol{\nabla}&=\varepsilon^{-4}\boldsymbol{\nabla}_z+\varepsilon^{-1}\boldsymbol{\nabla}_y+\boldsymbol{\nabla}_x,\\
\Delta&=\varepsilon^{-8}\Delta_z+\cdots.
\end{aligned}$$

After identification of the same order terms, we find

$$\boldsymbol{\nabla}_z p_m^0=\mathbf{0}, \tag{B4a}$$

$$\eta \Delta_z \boldsymbol{\nu}_m^1 - \boldsymbol{\nabla}_z p_m^3 - \boldsymbol{\nabla}_y p_m^0 = 0, \tag{B4b}$$

$$\boldsymbol{\nabla}_z \cdot \boldsymbol{\nu}_m^1 = 0, \tag{B4c}$$

$$j\omega\rho_m^0 + \rho_0 \boldsymbol{\nabla}_z \cdot \boldsymbol{\nu}_m^4 + \rho_0 \boldsymbol{\nabla}_y \cdot \boldsymbol{\nu}_m^1 = 0. \tag{B4d}$$

From (B4a), one gets that the pressure in the micropores does not depend on **z**, but it *a priori* depends on **x** and **y**:

$$p_m^0 = p_m^0(\mathbf{x},\mathbf{y}). \tag{B5}$$

Regarding Eq. (B4b), it must be underlined that this local Navier–Stokes equation does not involve the macroscopic pressure gradient, but $\boldsymbol{\nabla}_y p_m^0$.

Adding the motionless skeleton hypothesis ($\boldsymbol{\nu}_{m/\Gamma_{sm}}^1 = \mathbf{0}$), the system made of equations is classically solved using the associated variational formulation[16] and leads to the following solution:

$$\boldsymbol{\nu}_m^1 = -\frac{[\pi_m^{\nu}(\mathbf{z})]}{\eta} \boldsymbol{\nabla}_y p_m^0. \tag{B6}$$

In this equation tensor, the local $[\pi_m^{\nu}(\mathbf{z})]$ is purely real, because inertial forces are neglected in the frequency range under interest.

The following step consists in determining p_m^0, the pressure in the micropores. For this purpose we go to the mesoscopic scale.

Thermal effects are neglected (isothermal hypothesis), in the first place, so that

$$\frac{p_m^0}{P_0} = \frac{\rho_m^0}{\rho_0}. \tag{B7}$$

Integrating Eqs. (B4c) and (B7) over Ω_m yields

$$j\omega\phi_m \frac{p_m^0}{P_0} + \boldsymbol{\nabla}_y \cdot \langle \boldsymbol{\nu}_m^1 \rangle_{\Omega_m} = 0. \tag{B8}$$

Introducing Eqs. (B6) in (B8) gives

$$j\omega\phi_m \frac{p_m^0}{P_0} - \frac{\Pi_m(0)}{\eta} \Delta_y p_m^0 = 0, \quad \text{in } \Omega_m, \tag{B9}$$

where $\Pi_m(0)$ $[=(1/\Omega_m)\int_{\omega_{fm}} \pi_m^{\nu}\, d\Omega]$, is the static value of the permeability of the microporous medium. In this last equation Ω_m is supposed to be isotropic ($[\Pi_m(0)] = \Pi_m(0)\cdot[I]$).

2. In pores

Physics is not modified by the micropores, and we find that

$$p_p^0 = p_p^0(\mathbf{x}), \tag{B10a}$$

$$\boldsymbol{\nu}_p^0 = -\frac{\pi_p(\omega,\mathbf{y})}{\eta} \boldsymbol{\nabla}_x p_p^0, \tag{B10b}$$

$$j\omega\rho_p^0 + \rho_0 \boldsymbol{\nabla}_y \cdot \boldsymbol{\nu}_p^1 + \rho_0 \boldsymbol{\nabla}_x \cdot \boldsymbol{\nu}_p^0 = 0. \tag{B10c}$$

As results, pressure is uniform in pores (B10a) at the mesoscale, and the local flow law is given by the local dynamic Darcy law in pores (B10b).

Adding the boundary condition

$$p_m^0(\mathbf{x},\mathbf{y}) = p_p^0(\mathbf{x}), \quad \text{on } \Gamma_{sp}, \tag{B11}$$

and rewriting Eq. (B9), with $p^0 = p_m^0 - p_p^0$, one gets

$$\frac{P_0}{\phi_m} \frac{\Pi_m(0)}{\eta} \Delta_y p^0 - j\omega p = j\omega p_p^0,$$
$$p^0_{/\Gamma_{sp}} = 0. \tag{B12}$$

The so-posed problem looks like the diffusion problem of temperature in a ''single'' porosity medium, and the solution can be written as

$$p^0 = -j\omega \frac{\phi_m \eta}{P_0 \Pi_m(0)} d(\mathbf{y},\omega) p_p^0(\mathbf{x}). \tag{B13}$$

Hence, applying $\langle\cdot\rangle_{\Omega_{sp}} = (1/\Omega_p)\int_{\Omega_{sp}} \cdot d\Omega$ to Eq. (B8) yields

$$j\omega(1-\phi_p)\frac{\phi_m}{P_0} F_d(\omega) p_p^0 + \boldsymbol{\nabla}_y \cdot \langle \langle \boldsymbol{\nu}_m^1 \rangle_{\Omega_m} \rangle_{\Omega_{sp}} = 0, \tag{B14}$$

with

$$F_d(\omega) = \left(1 - j\omega \frac{\phi_m \eta}{P_0 \Pi_m(0)(1-\phi_p)} D(\omega)\right) \tag{B15}$$

and

$$D(\omega) = \frac{1}{\Omega_p} \int_{\Omega_{sp}} d(\mathbf{y},\omega)\, d\Omega. \tag{B16}$$

The macroscopic description is finally obtained, applying $\langle\cdot\rangle_{\Omega_{sp}} = (1/\Omega_p)\int_{\Omega_{fp}} \cdot d\Omega$ to Eq. (B10c) and using the continuity of flow on Γ_{sp},

$$\langle \boldsymbol{\nabla}_y \cdot \langle \boldsymbol{\nu}_m^1 \rangle_{\Omega_m} \rangle_{\Omega_{sp}} = -\langle \boldsymbol{\nabla}_y \cdot \boldsymbol{\nu}_p^1 \rangle_{\Omega_{fp}}. \tag{B17}$$

Assuming isothermal temperature in the pores, it gives

$$j\omega[\phi_p + \phi_m(1-\phi_p)F(\omega)]\frac{p_p^0}{P_0} + \boldsymbol{\nabla}_x \cdot \langle \boldsymbol{\nu}_p^0 \rangle_{\Omega_p} = 0. \tag{B18}$$

The macroscopic flow law been given by the generalized Darcy law in the porous material:

$$\langle \boldsymbol{\nu}_{db}^0 \rangle_{\Omega_p} = \langle \boldsymbol{\nu}_p^0 \rangle_{\Omega_p} = -\frac{[\Pi_p(\omega)]}{\eta} \boldsymbol{\nabla}_x p_p^0(\mathbf{x}). \tag{B19}$$

[1] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-filled-saturated porous solid—I. Low frequency range,'' J. Acoust. Soc. Am. **28**, 168–178 (1956).

[2] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-filled-saturated porous solid—II. Higher frequency range,'' J. Acoust. Soc. Am. **28**, 179–191 (1956).

[3] J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier Applied Science, London, 1993), p. 284.

[4] K. Attenborough, *Acoustical Characteristics of Porous Materials* (North-Holland, Amsterdam, 1982).

[5] R. Burridge and J. B. Keller, ''Poroelasticity equations derived from microstructure,'' J. Acoust. Soc. Am. **70**, 1140–1146 (1981).

[6] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949), p. 174.

[7] G. I. Barenblatt and I. P. Zheltov, ''On fundamental equations of flow of homogeneous liquids in naturally fractured rocks,'' Dokl. Akad. Nauk **132**, 545–548 (1960).

[8] J. R. Warren and P. J. Root, ''The behavior of naturally fractured reservoirs,'' Soc. Pet. Eng. J. 245–255 (1963).

[9] E. Sanchez-Palencia, *Non-Homogeneous Media and Vibration Theory*,

Lecture Notes in Physics, edited by J. Ehlers *et al.* (Springer-Verlag, Berlin, 1980), p. 398.
[10] J. L. Auriault and C. Boutin, ''Deformable media with double porosity– Quasi-statics. I: Coupling effects,'' Transp. Porous Media **7**, 63–82 (1992).
[11] J. L. Auriault and C. Boutin, ''Deformable media with double porosity— Quasi-statics. II: Memory effects,'' Transp. Porous Media **10**, 153–159 (1993).
[12] J. L. Auriault and C. Boutin, ''Deformable media with double porosity— III: Acoustics,'' Transp. Porous Media **14**, 143–162 (1994).
[13] P. Royer, ''Contribution de l'homogénéisation à l'étude de la filtration d'un gaz en milieu déformable à double porosité,'' (transl. ''Contribution of homogenization to the study of gas filtration through deformable medium with double porosity'') Ph.D. Thesis (Fourier, Grenoble, 1994), p. 422.
[14] C. Boutin, P. Royer, and J. L. Auriault, ''Acoustic absorption of porous surfacing with dual porosity,'' Int. J. Solids Struct. **35**, 4709–4737 (1998).
[15] X. Olny, ''Absorption acoustique des milieux poreux a simple et double porosité—Modélisation et validation expérimentale,'' (transl. ''Acoustic absorption of porous media with single and double porosity—Modeling and experimental validation''), Ph.D. thesis (ENTPE-INSA, Lyon, 1999), p. 281.
[16] J. L. Auriault, L. Borne, and R. Chambon, ''Dynamics of porous saturated media checking of the generalized law of Darcy,'' J. Acoust. Soc. Am. **77**, 1641–1650 (1985).
[17] D. Lafarge, P. Lemarinier, J. F. Allard, and V. Tarnow, ''Dynamic compressibility of air in porous structures at audible frequencies,'' J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[18] A. Bensoussan, J. L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures* (North-Holland, Amsterdam, 1978).
[19] D. L. Johnson, J. Koplik, and R. Dashen, ''Theory of dynamic permeability and tortuosity in fluid-saturated porous media,'' J. Fluid Mech. **1/6**, 379–402 (1987).
[20] V. Tarnow, ''Airflow resistivity of models of fibrous acoustic materials,'' J. Acoust. Soc. Am. **100**, 3706–3713 (1996).
[21] X. Olny and C. Boutin, ''The influence of pressure diffusion effects on absorbing properties of porous media with double porosity,'' in *Inter.noise 2000*, Nice, 2000.
[22] N. Atalla, F. Sgard, X. Olny, and R. Panneton, ''Acoustic absorption of macro-perforated porous materials,'' J. Sound Vib. **243**, 659–678 (2001).

Bifurcation of the Biot slow wave in a porous medium

Inna Edelman[a)]
Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany

(Received 28 May 2002; revised 11 April 2003; accepted 25 April 2003)

The propagation of the Biot slow wave in a fluid-saturated porous medium at low frequencies is investigated by asymptotic methods. It is proven that the Biot wave has a bifurcation behavior depending on its wave number. The bifurcation occurs in a neighborhood of the critical value k_{cr}, which depends on the permeability of a medium and the viscosity of a fluid. The $P2$ wave is fully attenuated if its wave number is smaller than k_{cr} and it becomes propagatory with wave numbers bigger than k_{cr}. Asymptotic formulas for the phase velocity and attenuation of the Biot wave are derived. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582443]

PACS numbers: 43.20.Bi, 43.20.Hq, 43.40.At [ANN]

I. INTRODUCTION

A fundamental theory for the propagation of elastic waves through an isotropic and macroscopically homogeneous fluid-saturated porous medium was developed by Biot.[1,2] Biot showed that one shear wave and two longitudinal waves, namely the fast ($P1$) and slow ($P2$) modes, can propagate in a fluid-saturated porous material. The shear wave and the longitudinal wave of the first kind ($P1$) predicted by the theory are similar to the corresponding waves in an ordinary isotropic elastic medium. The most interesting phenomenon, predicted by Biot, is the existence of a longitudinal wave of the second kind ($P2$), the so-called Biot slow wave. The main feature of this wave is that its phase velocity is almost always lower than both compressional wave velocities in a fluid and in a solid. The slow wave was observed experimentally at ultrasonic frequencies in *artificial rocks* made of sintered glass beads[3] and confirmed theoretically.[4] Later it was registered in *natural geomaterials*[5,6] also at ultrasonic frequencies.

In the low frequency range the Biot theory assumes that the slow wave is highly dispersive and strongly attenuated below some critical frequency, which depends on the pore size in the skeleton and the viscosity of the fluid. This critical frequency is typically around 1–10 kHz for water saturated porous materials of around 1 Darcy permeability.[7] The Biot slow wave is characterized by the out-of-phase motions of the solid skeleton and pore fluid. This relative motion is very sensitive to the viscosity of the fluid and the dynamic permeability of the porous medium.

Although Biot's theory has been thoroughly studied during last 40 years, the question of why the slow wave cannot be detected in low-permeability materials such as natural rocks is still open. The first attempt to answer this question was made in Ref. 8. The authors suggested that this lack of perceivable slow wave propagation is probably because of the clay particles which are deposited both within the pores and on the surfaces of the rock grains. They reduce the porosity and permeability of the rock and greatly increase the viscous losses. The latter results in the increasing attenuation of the slow wave and its complete disappearance.

The propagation of the slow wave in air-filled porous materials (porous ceramics of 2–70 Darcy and natural rocks of 200–700 mDarcy permeability) was investigated experimentally[9] between 10 and 500 kHz. The velocity and attenuation coefficient were measured as functions of frequency. It was shown that in the low frequency limit the phase velocity and attenuation of the Biot slow wave are essentially determined by the permeability of the porous material.

The purpose of this research is to prove analytically that the Biot slow wave possesses a bifurcation in a vicinity of some critical wave number which depends on the permeability of a medium and viscosity of a fluid. It is proven that the $P2$ wave is fully attenuated if its wave number is smaller than the critical value and it becomes propagatory with the wave numbers bigger than the critical one. The novelty of the work is an approach which results in the possibility to obtain an exact value for the critical point. In should be emphasized that we investigate the propagation of the acoustic waves through an infinite porous medium in the absence of external forces. In this case one must set the wave number k to be real and define the frequency $\omega=\omega(k)$, which can be complex, as a solution of the corresponding dispersion equation, i.e., we solve the initial value problem. This approach is adequate to the situation when we inspect which initial impulses, generated for instance as a result of explosions or earthquakes, can propagate and which can only diffuse. These theoretical results can give some hints on regions of nonpropagation of the Biot wave for the experimentalists performing field measurements.

The asymptotic analysis, presented in the paper, is based on the mathematical model of saturated poroelastic materials, proposed in Refs. 10–13. In order to investigate the velocities of the bulk waves, it suffices to consider its linear formulation, which evokes the classical Biot model. The comparison of the models is given in Ref. 14 as some explanations are given in the next section.

[a)]Permanent address: United Institute of Physics of the Earth, Russian Academy of Sciences, B. Gruzinskaya 10, Moscow 123242, Russia; electronic mail: edelman@wias-berlin.de

II. PROBLEM STATEMENT

A. Mathematical model

Consider the propagation of the bulk waves through an infinite space Ω occupied by a saturated porous medium. Balance equations describing the porous two-component medium have the following form[10–13] ($x \in \Omega,\ t \in [0,T]$).

Mass conservation equations,

$$\frac{\partial \rho^F}{\partial t} + \mathrm{div}(\rho^F \mathbf{v}^F) = 0, \quad \frac{\partial \rho^S}{\partial t} + \mathrm{div}(\rho^S \mathbf{v}^S) = 0. \tag{2.1}$$

Here ρ is the mass density, $\mathbf{v}$ is the velocity vector, and indices F and S indicate fluid or solid phases, respectively. The partial mass densities of the solid and fluid phases ρ^S and ρ^F are connected with the realistic (true) mass densities ρ^{SR} and ρ^{FR} by the relations: $\rho^S = (1-n)\rho^{SR}$, $\rho^F = n\rho^{FR}$, where n denotes the porosity (volume fraction of the fluid component).

Momentum conservation equations,

$$\begin{aligned} &\rho^F \frac{\partial \mathbf{v}^F}{\partial t} - \mathrm{div}\,\mathbf{T}^F + \pi(\mathbf{v}^F - \mathbf{v}^S) = 0, \\ &\rho^S \frac{\partial \mathbf{v}^S}{\partial t} - \mathrm{div}\,\mathbf{T}^S - \pi(\mathbf{v}^F - \mathbf{v}^S) = 0. \end{aligned} \tag{2.2}$$

Here $\mathbf{T}^F$ and $\mathbf{T}^S$ are the partial stress tensors.

It should be noted that in the model under research no explicit pore geometry is considered and, similar to Ref. 15, parameter π is referred to as the empirical constant to be determined. Biot,[1,2] considering the pore network as circular tubes, introduced the viscodynamic operator, which denotes deviation from the Poiseuille flow at high frequencies. Because of the fact that the main subject of the paper is the analysis of the behavior of the Biot slow wave in the low frequency range and the description of the bifurcation (transition) point between low and high frequency regimes, the effect of the frequency dependent (dynamic) permeability as well as effect of tortuosity, as discussed by Biot, are not taken into account (it is a well-known fact that at low frequencies only the static permeability plays an important role whereas tortuosity matters little). We presuppose $\pi = n_0 \mu^F/\mathcal{K}$, where n_0 is the initial value of porosity, μ^F is the viscosity of a liquid and $\mathcal{K}$ is the permeability of a porous medium.

Balance equation of porosity,

$$\frac{\partial n}{\partial t} + n_0\,\mathrm{div}(\mathbf{v}^F - \mathbf{v}^S) = -\frac{n-n_0}{\tau}, \tag{2.3}$$

where τ is the relaxation time of porosity, assumed to be constant. This equation describes the microscopic relaxation processes through the presence of the source term as well as the transport processes of porosity. This yields additional dissipation. Let us note that Biot[16] considered a balance equation of porosity similar to (2.3) without the right-hand side, i.e., without relaxation properties.

Constitutive relations for linear poroelastic materials,

$$\mathbf{T}^F = -p^F\mathbf{1} - \beta(n-n_0)\mathbf{1}, \quad p^F = p_0^F + \kappa(\rho^F - \rho_0^F), \tag{2.4}$$

$$\mathbf{T}^S = \mathbf{T}_0^S + \lambda^S I_1(\mathbf{e}^S)\mathbf{1} + 2\mu^S\mathbf{e}^S + \beta(n-n_0)\mathbf{1}. \tag{2.5}$$

Here p^F is the pore pressure, p_0^F and ρ_0^F are the initial values of pore pressure and fluid mass density, respectively, κ is the material parameter, assumed to be constant (in the special case, namely if the Biot parameter[2] $\alpha = n_0$, κ can be identified with the compressibility of the fluid; identification of the parameters of this model in terms of the Biot parameter[2] is discussed in Appendix). β is the coupling constant: corresponding terms in (2.4), (2.5) reflect the coupling between the solid and fluid phases due to the changes of porosity. $\mathbf{T}_0^S$ denotes a constant reference value of the partial stress tensor in the skeleton; λ^S and μ^S are the Lamé constants of the skeleton. $\mathbf{e}^S$ is the tensor of small deformations and $I_1(\mathbf{e}^S)$ is its first invariant. The parameters β and τ are experimental constants.

As it was demonstrated earlier (see explanations in Ref. 14), the coupling effects due to the changes of porosity influence velocities of $P1$ and $P2$ waves much less than those due to the relative motion of fluid and solid phases. Therefore, in the asymptotic analysis of wave propagation we can assume for simplicity that $\beta = 0$. In the Biot model, commonly used in the research of propagation of acoustic waves in a fluid-saturated porous medium, coupling between the skeleton and saturating liquid exists. However, one can also ignore this effect in the description of gross behavior of sound waves (see the Appendix).

The assumption $\beta = 0$ results in the fact that Eqs. (2.1), (2.2) can be solved independently of the equation for porosity (2.3) and, therefore, the relaxation time τ, appearing in (2.3), does not influence the results presented below. More accurate, τ does not enter the dispersion equation under research, since the dispersion equation is derived from (2.1), (2.2).

Moreover, the case $\beta = 0$ is considered deliberately: this simplification allows one to solve the problem separately from (2.3), which, in turn, leaves one the room for comparison of the results presented in the paper with those characteristic for the classical Biot model.

B. Dimensionless variables and parameters

Let us rewrite the system of equations (2.1)–(2.3) in a dimensionless form. For this purpose we introduce the following dimensionless variables and parameters:[14]

$$\hat{\rho}^F = \frac{\rho^F}{\rho_0^S}, \quad \hat{\rho}^S = \frac{\rho^S}{\rho_0^S}, \quad \hat{\mathbf{v}}^F = \frac{\mathbf{v}^F}{U_\parallel^S}, \quad \hat{\mathbf{v}}^S = \frac{\mathbf{v}^S}{U_\parallel^S},$$

where ρ_0^S is the initial value of the skeleton mass density and $U_\parallel^S = \sqrt{(\lambda^S + 2\mu^S)/\rho_0^S}$ is a velocity of a longitudinal wave in an unbounded elastic medium. Also one has

$$\hat{x} = \frac{x}{U_\parallel^S \tau}, \quad \hat{t} = \frac{t}{\tau}, \quad \hat{p}^F = \frac{p^F}{\rho_0^S (U_\parallel^S)^2}, \quad \hat{\kappa} = \frac{\kappa}{(U_\parallel^S)^2},$$

$$\hat{\pi} = \frac{\pi\tau}{\rho_0^S}, \quad \hat{\beta} = \frac{\beta}{\rho_0^S (U_\parallel^S)^2}, \quad \hat{\lambda}^S = \frac{\lambda^S}{\rho_0^S (U_\parallel^S)^2}, \quad \hat{\mu}^S = \frac{\mu^S}{\rho_0^S (U_\parallel^S)^2}.$$

After the change of variables and parameters, the original system (2.1)–(2.3) keeps its form except for the right-hand

side in the balance equation of porosity. One gets there $-(n-n_0)$. We will subsequently drop the overcaret symbol characterizing dimensionless quantities.

The introduction of the dimensionless variables is stipulated only by the reasons of convenience (the resulting formulas are shorter and more readable). Obviously, the dimensionless variables are dependent on the parameter τ, however this dependence is eliminated when returning to the physical variables.

C. Dispersion equation for the bulk waves

Let us investigate the propagation of the bulk waves through an infinite porous medium. We will only study the propagation of the longitudinal waves. Therefore it suffices to consider the 1D problem. In this case the system (2.1)–(2.3) reads (it is assumed $\beta=0$):

$$\begin{aligned}
&\frac{\partial \rho^F}{\partial t}+\frac{\partial}{\partial x}(\rho^F v^F)=0, \quad \frac{\partial \rho^S}{\partial t}+\frac{\partial}{\partial x}(\rho^S v^S)=0,\\
&\rho^F\frac{\partial v^F}{\partial t}+\kappa\frac{\partial \rho^F}{\partial x}+\pi(v^F-v^S)=0,\\
&\rho^S\frac{\partial v^S}{\partial t}-(\lambda^S+2\mu^S)\frac{\partial e^S}{\partial x}-\pi(v^F-v^S)=0, \quad \frac{\partial e^S}{\partial t}=\frac{\partial v^S}{\partial x},\\
&\frac{\partial n}{\partial t}+n_0\frac{\partial}{\partial x}(v^F-v^S)=-(n-n_0).
\end{aligned} \tag{2.6}$$

Consider the propagation of the harmonic waves whose frequency is ω and wave number is k. Below we use the following dimensionless parameters: $\hat{\omega}=\omega\tau$ and $\hat{k}=kU_{\|}^S\tau$ (the overcaret symbol is again omitted in further consideration). Substituting solutions in the form

$$\begin{aligned}
&\rho^F-\rho_0^F=R^F\exp(i(kx-\omega t)),\\
&\rho^S-\rho_0^S=R^S\exp(i(kx-\omega t)),\\
&v^F=V^F\exp(i(kx-\omega t)),\\
&v^S=V^S\exp(i(kx-\omega t)),\\
&e^S=E\exp(i(kx-\omega t)),\\
&n-n_0=D\exp(i(kx-\omega t))
\end{aligned} \tag{2.7}$$

into equation system (2.6), one gets the system of algebraic equations for the unknown amplitudes.[18] Requesting that the determinant of this system must vanish yields the dispersion equation for the longitudinal waves

$$\mathcal{F}(k,\omega)=0, \tag{2.8}$$

where

$$\begin{aligned}
\mathcal{F}(k,\omega)=&r(\omega^2-c_f^2k^2)(\omega^2-k^2)+i\omega\pi((1+r)\omega^2\\
&-k^2(1+rc_f^2)),
\end{aligned} \tag{2.9}$$

$r=\rho_0^F/\rho_0^S$, $c_f=U^F/U_{\|}^S$, and $U^F=\sqrt{\kappa}$.

As previously noted, similar to Refs. 14, 17, and 18 we consider the propagation of the acoustic waves through an infinite domain in the absence of external forces, so that the solutions for the system (2.6) are defined uniquely by the Cauchy data (initial value problem). In this case the frequency ω is defined by the dispersion equation (2.8) as a function of the real wave number $k\in R^1$. Thus, Re ω/k defines the phase velocity of a wave and Im ω gives its attenuation.

Our goal is to prove that the solution $\omega_{P2}(k)$ of the dispersion equation (2.8), corresponding to the Biot slow wave, possesses a bifurcation. It takes place at some critical point k_{cr} (bifurcation point), in a small neighborhood of which the solution of Eq. (2.8) splits into several branches.

III. BIFURCATION OF THE BIOT SLOW WAVE

Let us rewrite Eq. (2.8) as

$$r(\tilde{\omega}^2-c_f^2)(\tilde{\omega}^2-1)+i\tilde{\omega}\frac{1}{\tilde{k}}((1+r)\tilde{\omega}^2-(1+rc_f^2))=0, \tag{3.1}$$

where $\tilde{\omega}=\omega/k$ and $\tilde{k}=k/\pi$. Obviously, for the case $k\gg 1$ (high frequency range), Eq. (3.1) has the roots (note that here $1/\tilde{k}\ll 1$ is assumed to be a small parameter)

$$\tilde{\omega}_{P1}=\pm 1-\frac{i}{2}\frac{1}{\tilde{k}}\mp\frac{4+r-rc_f^2}{8r(1-c_f^2)}\frac{1}{\tilde{k}^2}+O\left(\frac{1}{\tilde{k}^3}\right) \tag{3.2}$$

and

$$\tilde{\omega}_{P2}=\pm c_f-\frac{i}{2r}\frac{1}{\tilde{k}}-\frac{1-c_f^2(1+4r)}{8r^2(1-c_f^2)(\pm c_f)}\frac{1}{\tilde{k}^2}+O\left(\frac{1}{\tilde{k}^3}\right), \tag{3.3}$$

which define the velocities and attenuations of forward and backward directed longitudinal waves of the first and second kinds, respectively. It is evident that in the high frequency limit, the phase velocities of $P1$ and $P2$ waves do not depend on the frequency ω.

Next let us consider the low frequency range, when $k\ll 1$ and, consequently, $\tilde{k}\ll 1$. The solutions of Eq. (3.1) are sought in the following form:

$$\tilde{\omega}=\tilde{\omega}_0+\tilde{k}\tilde{\omega}_1+\tilde{k}^2\tilde{\omega}_2+\cdots. \tag{3.4}$$

For the longitudinal $P1$ wave of forward and backward directions one obtains

$$\begin{aligned}
\tilde{\omega}_{P1}=&\pm\sqrt{\frac{1+rc_f^2}{1+r}}-\tilde{k}\frac{ir(1-c_f^2)^2}{2(1+rc_f^2)(1+r)^2}\\
&\pm\tilde{k}^2\sqrt{\frac{1+r}{1+rc_f^2}}\frac{r^2(1-c_f^2)^3(2(1-r)(1+rc_f^2)+1-c_f^2)}{8(1+r)^4(1+rc_f^2)^2}\\
&+O(\tilde{k}^3).
\end{aligned} \tag{3.5}$$

However the $P2$ wave construction of the asymptotic solution for the corresponding root of (3.1) is much more complicated. We prove later on that there exists some critical value of wave number k_{cr}, below which the longitudinal wave of the second kind is not propagatory. Thus, the asymptotic expansion of the corresponding root of (3.1) has a different structure depending on whether the wave number of the $P2$ wave is smaller or bigger than its critical value k_{cr}.

Substitution of (3.4) into (3.1) yields for the forward directed $P2$ wave,

$$\tilde{\omega}^f_{P2}=-i\frac{rc_f^2}{1+rc_f^2}\tilde{k}-i\frac{r^3c_f^4(1+rc_f^4)}{(1+rc_f^2)^4}\tilde{k}^3+O(\tilde{k}^4). \tag{3.6}$$

The solution for the backward directed $P2$ wave is sought in the form

$$\tilde{\omega}=\frac{1}{\tilde{k}}\tilde{\omega}_0+\tilde{\omega}_1+\tilde{k}\tilde{\omega}_2+\cdots \tag{3.7}$$

and it leads to the expansion

$$\tilde{\omega}^b_{P2}=-i\frac{1+r}{r}\frac{1}{\tilde{k}}+i\frac{r(r+c_f)}{(1+r)^2}\tilde{k}+O(\tilde{k}^2). \tag{3.8}$$

Obviously, the expansions (3.6) and (3.8) consist of the imaginary terms only. The latter means that the phase velocity of the $P2$ wave is equal to zero, i.e., the wave is not propagatory (it is fully attenuated). However, these expansions are valid only if the wave number k is smaller than some critical value $k_{\rm cr}$. In other words, there exists a bifurcation point $k_{\rm cr}$ in a small neighborhood of which the corresponding solution of (2.8) splits into several branches. Let us prove this statement. Consider the dispersion equation (2.8). It is easy to see that the exact solution for the $P2$ wave is given by

$$k^2=\frac{1}{2rc_f^2}(r\omega^2(1+c_f^2)+i\pi\omega(1+rc_f^2)+\sqrt{r^2\omega^4(1-c_f^2)^2-\pi^2\omega^2(1+rc_f^2)^2+2ir\pi\omega^3(1-c_f^2)(1-rc_f^2)}). \tag{3.9}$$

Proposition: There exists some critical value of the wave number $k_{\rm cr}\in R^+$ such that (a) if $0<k<k_{\rm cr}$ then Eq. (3.9) has two pure imaginary roots $\omega_1(k)$ and $\omega_2(k)$, $\mathrm{Re}\,\omega_j(k)=0$, $j=1,2$; (b) if $k=k_{\rm cr}$ then Eq. (3.9) has one multiple pure imaginary root, i.e., $\omega_1(k)=\omega_2(k)$, $\mathrm{Re}\,\omega_j(k)=0$, $j=1,2$; (c) if $k>k_{\rm cr}$ then Eq. (3.9) has no pure imaginary roots.

Proof: Applying the change $\omega=-i\pi\Omega$, $\Omega\geqslant 0$, Eq. (3.9) can be rewritten as

$$F_1(\Omega)=F_2(\Omega), \tag{3.10}$$

where

$$F_1(\Omega)=\Omega\sqrt{\Omega^2r^2(1-c_f^2)^2-2r\Omega(1-c_f^2)(1-rc_f^2)+(1+rc_f^2)^2}, \tag{3.11}$$

$$F_2(\Omega)=2rc_f^2\tilde{k}^2+\Omega^2r(1+c_f^2)-\Omega(1+rc_f^2), \tag{3.12}$$

and, as above, $\tilde{k}=k/\pi$. It should be noted here that the function under the square root $g(\Omega)=\Omega^2r^2(1-c_f^2)^2-2r\Omega(1-c_f^2)(1-rc_f^2)+(1+rc_f^2)^2$ is always positive. Consider Eq. (3.10). First let us investigate the behavior of the functions $F_1(\Omega)$, $F_2(\Omega)$ as $\Omega\to\infty$. Obviously,

$$\frac{F_1(\Omega)}{\Omega^2}\sim r(1-c_f^2)\quad\text{and}\quad\frac{F_2(\Omega)}{\Omega^2}\sim r(1+c_f^2), \tag{3.13}$$

i.e., $F_2(\Omega)$ is steeper than $F_1(\Omega)$. Consequently, if $\tilde{k}=0$ then the function $(F_2-F_1)(\Omega)$ has two real roots: $\Omega=0$ and some Ω_*, so that $(F_2-F_1)(\Omega)<0$ in $(0,\Omega_*)$.

Next we calculate the stationary points for $F_1(\Omega)$ and $F_2(\Omega)$ and the inflection points for $F_1(\Omega)$. One can easily verify that the function $F_1(\Omega)$ has two stationary points, namely $\Omega_1^{(1)}\approx(1+(1+7r)c_f^2)/(2r)$ and $\Omega_1^{(2)}\approx(1+(1-5r)c_f^2)/r$ and the function $F_2(\Omega)$ has one stationary point $\Omega_2=(1+rc_f^2)/(2r(1+c_f^2))$ such that

$$\Omega_2<\Omega_1^{(1)}<\Omega_1^{(2)}. \tag{3.14}$$

Function $F_1(\Omega)$ has a unique inflection point

$$\Omega_{\rm inf}\approx\frac{1-rc_f^2-\sqrt[3]{2}\sqrt[3]{r}\sqrt[3]{c_f^2}(1-\sqrt[3]{2}\sqrt[3]{r}\sqrt[3]{c_f^2})}{r(1-c_f^2)} \tag{3.15}$$

and it is known that $\Omega_1^{(1)}<\Omega_{\rm inf}<\Omega_1^{(2)}$ as well as that $F_1(\Omega)$ is concave if $\Omega<\Omega_i$ and $F_1(\Omega)$ is convex if $\Omega>\Omega_i$. The preceding analysis allows us to conclude that there exists a unique point of tangency of the functions $F_1(\Omega)$ and $F_2(\Omega)$ (see Fig. 1). Thus, using $F_1'(\Omega)=F_2'(\Omega)$, one can define the point of tangency, i.e., the critical value $\Omega_{\rm cr}$:

$$\Omega_{\rm cr}\approx\frac{1}{2r}+2c_f^2(1+3rc_f^2-2c_f^2), \tag{3.16}$$

which is positive by virtue of physical sense. The corresponding critical value of the wave number is defined from Eq. (3.10) and it is given by

$$\tilde{k}_{\rm cr}\approx c_f\left(1+\frac{1}{2rc_f^2}\right),\quad\text{i.e.,}\quad k_{\rm cr}\approx c_f\left(1+\frac{1}{2rc_f^2}\right)\pi. \tag{3.17}$$

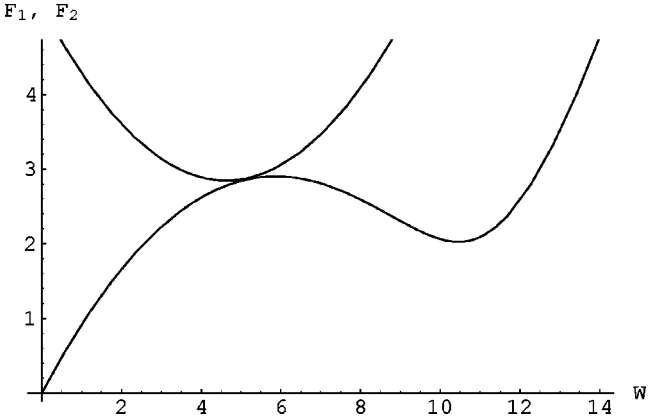


FIG. 1. Numerical example: $r=0.1$, $c_f=0.32$, $\tilde{k}=\tilde{k}_{\rm cr}=15.9$.

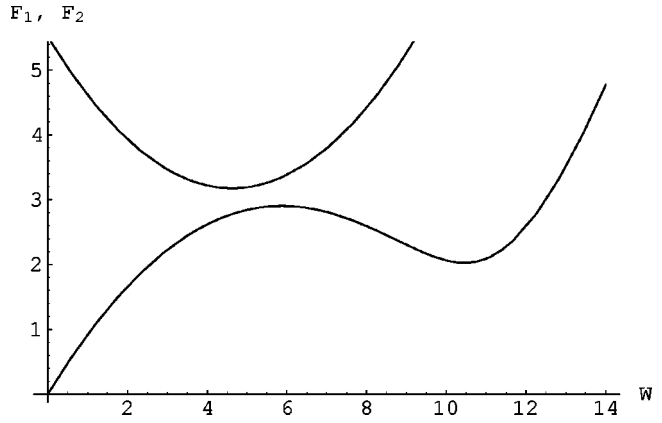


FIG. 2. Numerical example: $r=0.1$, $c_f=0.32$, $\tilde{k}=17.5>\tilde{k}_{\rm cr}$.

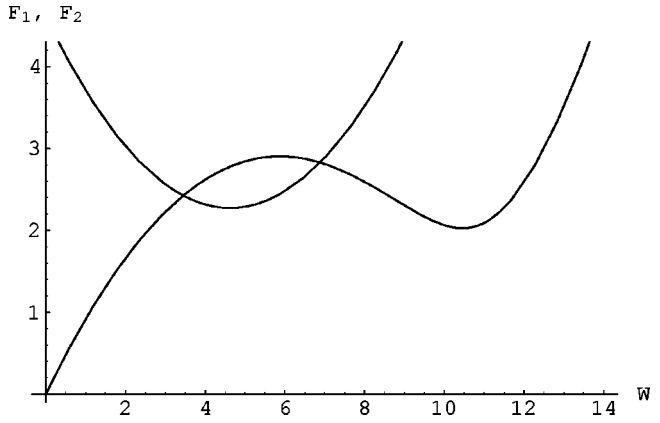


FIG. 3. Numerical example: $r=0.1$, $c_f=0.32$, $\tilde{k}=15.5<\tilde{k}_{\rm cr}$.

Therefore, it was proven that there exists some critical real value $k_{\rm cr}$ for which Eq. (3.9) has one multiple imaginary root $\omega_{\rm cr}=-i\pi\Omega_{\rm cr}$.

Next we prove that if $k<k_{\rm cr}$ then Eq. (3.9) has two pure imaginary roots and if $k>k_{\rm cr}$ then Eq. (3.9) has no imaginary roots. Consider the expansions

$$k=k_{\rm cr}(1\pm\epsilon k_1\pm\epsilon^2 k_2+\cdots)$$

and

$$\omega=\omega_{\rm cr}+\epsilon\omega_1+\epsilon^2\omega_2+\cdots, \tag{3.18}$$

where ϵ is a small parameter. Note that here any ϵ can be chosen. For example, one may set $\epsilon=c_f^{\delta}$, $0<\delta\leqslant 1$. Substitution of (3.18) into (3.9) yields the bifurcation equation. From its $O(\epsilon)$ approximation it follows that $k_1=0$. From the next $O(\epsilon^2)$ approximation one has

$$\pm k_2=\frac{1}{4}\frac{\omega_1^2}{k_{\rm cr}^2}\mathcal{A} \tag{3.19}$$

with

$$\mathcal{A}=\frac{1+c_f^2}{c_f^2}+\frac{1-c_f^2}{c_f^2 g(\Omega_{\rm cr})\sqrt{g(\Omega_{\rm cr})}}(-r^3(1-c_f^2)^3\Omega_{\rm cr}^3 +3r^2(1-c_f^2)^2(1-rc_f^2)\Omega_{\rm cr}^2-3r(1-c_f^2)(1+r^2c_f^4)\Omega_{\rm cr} +(1-rc_f^2)(1+rc_f^2)^2)>0. \tag{3.20}$$

It is obvious that for the given $k_2>0$ there are two real solutions for ω_1 for Eq. (3.19) if the plus sign is chosen on the left-hand side. The letter means that we consider the expansion $k=k_{\rm cr}(1+\epsilon^2 k_2+\cdots)$ and $k>k_{\rm cr}$. Consequently, Eq. (3.10) has no solution (see Fig. 2).

Vice versa, if $k=k_{\rm cr}(1-\epsilon^2 k_2+\cdots)<k_{\rm cr}$ then for the given $k_2>0$ Eq. (3.19) as well as Eq. (3.9) have two imaginary roots (see Fig. 3). Thus, the proposition was proven.

Therefore we conclude that the $P2$ wave is not propagatory (fully attenuated) if its wave number is smaller than the critical value $k_{\rm cr}$. Otherwise, the Biot wave begins to emerge with the frequency

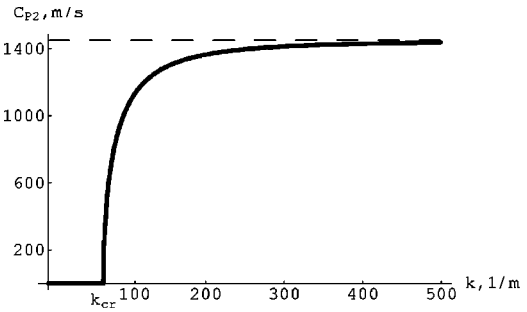
$$\omega_{P2}=\omega_{\rm cr}+\epsilon\omega_1+\pi O(\epsilon^2),\quad \omega_1=2k_{\rm cr}\sqrt{k_2/\mathcal{A}}. \tag{3.21}$$

Consequently, the phase velocity of the forward directed $P2$ wave in the low frequency range (more precisely, in a small neighborhood of the bifurcation point) is given by the formula

$$c_{P2}=\mathrm{Re}(\omega_{P2})/k, \tag{3.22}$$

where $k=k_{\rm cr}(1+\epsilon^2 k_2)+\pi O(\epsilon^3)$ and $k_2=O(1)$ with respect to the small parameter ϵ. Obviously, the Biot wave is strongly attenuated at low frequencies [the leading term in the expansion (3.21) is complex]. In Fig. 4 the phase velocity of the $P2$ wave and the real part of its frequency are evaluated for water-saturated sandstone (see example below).

It should be emphasized once again that the assumption $\beta=0$ leads to the fact that the results being presented do not depend on the parameter τ appearing in Eq. (2.3). For example, the dimensionless formula (3.17) reads in physical variables



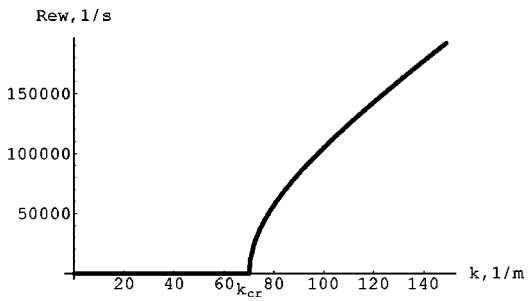


FIG. 4. Phase velocity (left-handside) and real part of frequency (right-handside) of the P2 wave, evaluated for water-saturated sandstone: $k_{\rm cr}\approx 70.22\ \mathrm{m}^{-1}$.

$$k_{\rm cr} \approx c_f \left(1 + \frac{1}{2rc_f^2} \right) \frac{\pi}{\rho_0^S U_{\|}^S}.$$

Analogously, the critical frequency evaluated in dimensional variables has the form

$$\omega_{\rm cr} = -\frac{i\pi}{\rho_0^S} \Omega_{\rm cr},$$

where $\Omega_{\rm cr}$ is defined by (3.16).

IV. NUMERICAL EXAMPLE AND COMPARISON WITH THE BIOT THEORY

The second formula in (3.17) shows clearly that the critical wave number $k_{\rm cr}$ is directly proportional to the parameter π. The corresponding critical wavelength's $\lambda_{\rm cr}$ dependence on π is through an inverse proportionality, i.e., through a direct proportionality with respect to the permeability $\mathcal{K}$. As it is proven above, the Biot slow wave is propagatory if its wave number $k > k_{\rm cr}$, in other words if its wavelength $\lambda < \lambda_{\rm cr}$.

To obtain estimates of the critical wavelength let us consider an example of practical importance, namely water-saturated sandstone with the following typical values of parameters:[19] initial porosity, $n_0 = 0.2$; initial value of the partial mass density of the fluid phase, $\rho_0^F = 0.2 \times 10^3$ kg/m^3. Note, that $\rho_0^F = n_0 \rho_0^{FR}$, where the initial value of the real mass density of the fluid phase $\rho_0^{FR} = 10^3$ kg/m^3; initial value of the partial mass density of the solid phase, $\rho_0^S = 2.0 \times 10^3$ kg/m^3. Note, that $\rho_0^S = (1 - n_0)\rho_0^{SR}$, where the initial value of the real mass density of the solid phase $\rho_0^{SR} = 2.65 \times 10^3$ kg/m^3; velocity of the compressional wave in the fluid, $U^F = 1450$ m/s; velocity of the longitudinal wave in the skeleton, $U_{\|}^S = 4450$ m/s; $\pi = n_0 \mu^f / \mathcal{K}$, where the viscosity of water $\mu^F = 10^{-3}$ kg/ms and the permeability of a porous medium $\mathcal{K} = 1000$ mDarcy, i.e., $\pi \approx 4 \times 10^7$ kg/m^{3}s.

Let us note that the parameter τ is absent in this list as one does not need it for the construction of the solutions of the dispersion equation. The dispersion equation is derived from the initial equations (2.1), (2.2) (prior to the introduction of the dimensionless variables) and, therefore τ does not appear in it.

The real part of frequency [numerical solution $\omega_{P2} = \omega_{P2}(k)$ of the dispersion equation, corresponding to the slow wave] and the phase velocity (numerical solution $\mathrm{Re}\, \omega_{P2}/k$) of the $P2$ wave, evaluated for the above parameters, are given in Fig. 4.

Evidently, the critical wave number is rather large: $k_{\rm cr} \approx 70.22$ m^{-1}. The pertinent critical wavelength $\lambda_{\rm cr} \approx 8.9$ cm. Thus, one can conclude that the Biot slow wave becomes propagatory with a short enough wavelength. The following observation can also be made: while the real part of frequency and, consequently, the velocity of the $P2$ wave are equal to zero at the critical point, they increase very rapidly in its small neighborhood $k > k_{\rm cr}$. One can see [and estimate from (3.21)], that in the ϵ-vicinity of the critical wave number real part of frequency grows so quickly, that it almost instantly achieves the value of approximately 2 kHz. This example allows us to draw an inference that the Biot slow wave cannot be detected in the low frequency range of interest in seismology (1–100 Hz). Moreover, a direct dependence of the critical wavelength $\lambda_{\rm cr}$ on the permeability $\mathcal{K}$, as it follows from (3.17), implies that in low permeable geomaterials the propagatory $P2$ modes have a relatively short wavelength. This deduction seems to be in agreement with the known experimental facts that in those materials low frequency regime extends to frequencies up to 1 MHz.

One can finally recall, that the Biot critical frequency is typically around 1–10 kHz for water saturated porous materials of around 1 Darcy permeability.[7] The above estimate of the critical wave number for the same medium reveals that both the classical Biot model and the model under research as presented lead to qualitatively and quantitatively similar results.

V. CONCLUSIONS

The results presented in this paper concern the propagation of the Biot slow wave in a fluid-saturated porous medium at low frequencies. For the first time it is proven analytically that the longitudinal wave of the second kind is fully attenuated if its wave number is lower than some critical value. This critical wave number is a bifurcation point, above which the longitudinal wave of the second kind becomes propagatory with the phase velocity given by (3.22). The bifurcation point describes the transition from low to high frequency regimes of propagation of the $P2$ mode. The asymptotic formulas (3.17) demonstrate that the Biot wave behavior is dominated by the permeability of a medium. The estimates of the critical values for wave number and wavelength allow one to suppose that in fluid-filled geological materials of low permeability the Biot slow wave does not propagate at seismic frequencies. Furthermore, the theoretical results presented in the paper might be helpful to experimentalists in the interpretation of field data. The foregoing analysis shows that the long wavelength $P2$ modes, generated for instance as a result of explosions or earthquakes, are not only strongly attenuated, as the Biot theory assumes, but may not be propagatory at all in the earth or porous rocks. It is extremely unlikely that the propagating slow waves will ever be directly observed in a field experiment in the relevant frequency range (10–1000 kHz).

ACKNOWLEDGMENTS

The author thanks Professors E. Radkevich (Moscow State University) and K. Wilmanski (WIAS) for helpful discussions. Support by the Alexander von Humboldt Foundation is gratefully acknowledged. The author also thanks Andrew N. Norris for a thorough review of this paper.

APPENDIX

First let us clarify the correspondence between the parameters of the model under research and those of the classical Biot model. In the Biot model the total stresses have the form (in the notations of Ref. 2):

$$\tau_{ij}=\lambda_c e\,\delta_{ij}+2\mu e_{ij}-\alpha M\zeta\,\delta_{ij}, \tag{A1}$$

where δ_{ij} is the Kronecker delta. Taking into account that $\zeta=n_0(e-\epsilon)$, one gets

$$\tau_{ij}=(\lambda_c-\alpha n_0 M)e\,\delta_{ij}+2\mu e_{ij}+\alpha n_0 M\epsilon\,\delta_{ij}. \tag{A2}$$

The fluid pressure is defined as

$$p_f=-\alpha Me+M\zeta=(-\alpha+n_0)Me-n_0M\epsilon. \tag{A3}$$

In the model under research the total stresses are of the form

$$T_{ij}=\lambda^S e\ \delta_{ij}+2\mu^S e_{ij}-\kappa(\rho^F-\rho_0^F)\delta_{ij}. \tag{A4}$$

Introducing macroscopic fluid strain ϵ as

$$\epsilon=-\frac{\rho^F-\rho_0^F}{\rho_0^F} \tag{A5}$$

one has

$$T_{ij}=\lambda^S e\,\delta_{ij}+2\mu^S e_{ij}+\kappa\rho_0^F\epsilon\,\delta_{ij}. \tag{A6}$$

The fluid pressure is defined by the following formula (as before we assume $\beta=0$)

$$p_f=\frac{p^F}{n_0}=-\frac{\rho_0^F\kappa}{n_0}\epsilon. \tag{A7}$$

Comparing the expressions (A2), (A6) and (A3), (A7) one can conclude that

$$\lambda^S=\lambda_c-\alpha n_0 M,\quad \mu^S=\mu, \tag{A8}$$

$$-\alpha+n_0=0,\quad \kappa\rho_0^F=n_0^2M,\quad \kappa\rho_0^F=\alpha n_0 M. \tag{A9}$$

From the last equation in (A9) it follows that the parameter κ is related to the Biot parameters α and M by $\rho^{FR}\kappa=\alpha M$. The first equation in (A9) implies $\alpha=n_0$. The latter means that the second and third equations are identical. It is easy to see also that in case $\alpha=n_0$ the parameter κ is the compressibility coefficient of the fluid. The definitions of the Lamé constants (A8) coincide completely with those given by Biot.[2] Let us note that, as Biot proved,[20] in general $n_0\leqslant\alpha\leqslant 1$. Later it has been shown[15,21] that $n_0<\alpha<1$. However, in materials with stiff pore space (e.g., small fractured limestones, carbonate rocks, bones) the coefficient α approaches its lower bound and we can assume $\alpha\approx n_0$.

As a result, Eqs. (2.1), (2.2) as presented are a particular case of the Biot model with $\alpha=n_0$.

The difference in the models is related to Eq. (2.3). However, as it has been mentioned early on, the assumption $\beta=0$ abolishes the dependence of the results on the relaxation time τ. Furthermore, if the relaxation of porosity is ignored in (2.3), i.e., if Eq. (2.3), similar to Ref. 16, is considered without the right-hand side, then the model (2.1)–(2.3) is a special case of the Biot model providing $\alpha=n_0$ and $\beta=0$. The general case of the model and the discussion of the definition of experimental parameters β and τ can be found elsewhere.[10–13] Here we investigate premeditatedly the simplified case $\beta=0$ with a view to present the model under research as close to the classical Biot model as possible.

A final remark concerns the Biot coupling between the skeleton and saturating liquid in Eqs. (A2), (A3), where the solid strain and the fluid strain are defined as $e=\operatorname{div}\mathbf{u}$ and $\epsilon=\operatorname{div}\mathbf{U}$. Here $\mathbf{u}$ and $\mathbf{U}$ denote the displacements of the skeleton and the fluid displacements, observed from the surface of the frame. It should be noted that both the classical Biot model and the model under research are based on the approach which consists of deliberately ignoring the *microscopic* level and assuming that the concepts and principles of continuum mechanics can be applied to measurable *macroscopic* values. However, the fluid displacements $\mathbf{U}$, introduced in the Biot model, are not identified with the macroscopic displacement[19] and, consequently, the Biot fluid strain has no macroscopic meaning. Macroscopic fluid strain is defined by (A5). Thus, if we would take into account additional coupling, required in the Biot model, the stress tensors (2.4), (2.5) would include the terms $Q\,I_1(\mathbf{e}^S)$ and $Q[(\rho^F-\rho_0^F)/\rho_0^F]$, respectively, where $Q=n_0M(\alpha-n_0)$ is the Biot coupling constant. Let us estimate an influence of the Biot coupling. This assessment is easily done by comparing the terms entering (2.4), (2.5). The coefficient Q has an order of magnitude of the pore pressure,[1,20] i.e., 10^5 Pa in soils and rocks. The elastic constants λ^S, μ^S, and $\kappa\rho_0^F$ are of the order 10^8-10^9 Pa[19]. Therefore, assuming the coupling effect through Q is small enough, we may ignore it in the asymptotic analysis of wave propagation.

[1] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solids I. Low frequency range & II. High frequency range," J. Acoust. Soc. Am. **28**, 168–186 (1956).
[2] M. A. Biot, "Mechanics of deformations and acoustic propagation in porous media," J. Appl. Phys. **33**, 1482–1498 (1962).
[3] T. J. Plona, "Observation of a second bulk compressional wave in a porous medium at ultrasonic frequencies," Appl. Phys. Lett. **36**, 259–261 (1980).
[4] J. G. Berryman, "Confirmation of Biot's theory," Appl. Phys. Lett. **37**, 382–384 (1980).
[5] K. Nakagawa, K. Soga, and J. K. Mitchell, "Observation of Biot compressional wave of the second kind in granular soils," Geotechnique **47**, 133–147 (1997).
[6] O. Kelder and D. M. J. Smeulders, "Observation of the Biot slow wave in water-saturated Nivelsteiner sandstone," Geophysics **62**, 1794–1796 (1997).
[7] P. B. Nagy, "Acoustics and ultrasonics," in *Experimental Methods in the Physical Sciences* (Academic New York, 1999), pp. 161–221.
[8] T. Klimentos and C. McCan, "Why is the Biot slow compressional wave not observed in real rocks?," Geophysics **53**, 1605–1609 (1988).
[9] P. B. Nagy, "Slow wave propagation in air-filled permeable solids," J. Acoust. Soc. Am. **93**, 3224–3234 (1993).
[10] K. Wilmanski, "Lagrangean model of two-phase porous material," J. Non-Equilib. Thermodyn. **20**, 50–77 (1995).
[11] K. Wilmanski, "Porous media at finite strains. The new model with the balance equation for porosity," Arch. Mech. **48**, 591–628 (1996).
[12] K. Wilmanski, "A thermodynamic model of compressible porous materials with the balance equations of porosity," Transp. Porous Media **32**, 21–47 (1998).
[13] K. Wilmanski, *Thermomechanics of Continua* (Springer-Verlag, Berlin, 1998).
[14] I. Edelman and K. Wilmanski, "Asymptotic analysis of surface waves at vacuum/porous medium and liquid/porous medium interfaces," Continuum Mech. Thermodyn. **14**, 25–44 (2002).
[15] A. N. Norris, "Stoneley-wave attenuation and dispersion in permeable formations," Geophysics **54**, 330–341 (1989).
[16] M. A. Biot, "Generalized theory of three dimensional consolidation," J. Appl. Phys. **12**, 155–161 (1941).
[17] I. Edelman, "Waves on boundaries of porous media," Phys. Dokl. **46**, 517–521 (2001).

[18] I. Edelman, ''On the bifurcation of the Biot slow wave in a porous medium'' (unpublished). WIAS—Preprints are easily available from the webpage http://www.wias-berlin.de/publications/preprints as pdf or ps-files.
[19] T. Bourbie, O. Coussy, and B. Zinszner, *Acoustics of Porous Media* (Technip, Paris, 1987).
[20] M. Biot and D. Willis, ''The elastic coefficients of the theory of consolidation,'' J. Appl. Mech. **24**, 594–601 (1957).
[21] R. W. Zimmerman, W. H. Somerton, and M. S. King, ''Compressibility of porous rocks,'' J. Geophys. Res., [Solid Earth Planets] **91**, 12765–12777 (1986).

Empirical model of the acoustic impedance of a circular orifice in grazing mean flow

Seong-Hyun Lee and Jeong-Guon Ih[a)]
Center for Noise and Vibration Control, Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Science Town, Taejon 305-701, Korea

(Received 15 June 2002; revised 30 March 2003; accepted 21 April 2003)

Although there are many analytical and empirical models for orifice impedance, the predicted acoustical performance when adopting any one of them sometimes shows a large discrepancy with the measured result in some cases. In order to obtain a new practical and precise empirical impedance model under grazing flow conditions, the acoustic impedance of circular orifices has been measured with a variation of the involved parameters under very carefully tested and controlled measurement conditions. The parameters involved in determining the acoustic impedance of an orifice are comprised of the orifice diameter, orifice thickness, perforation ratio, mean flow velocity, and frequency. The range of involved parameters is chosen to cover the practical data span of perforates in typical exhaust systems of internal combustion engines. The empirical impedance model is obtained by using nonlinear regression analysis of the various results of the parametric tests. The proposed empirical model of orifice impedance, with a very high correlation coefficient, is applied to the prediction of the transmission loss of concentric resonators, which have geometric configurations typical of acoustically short and long through-flow resonators. By comparing the measured and predicted results, in which the predictions are made by employing many previous orifice impedance models as well as the present model, it is confirmed that the proposed orifice impedance model yields the most accurate prediction among all other existing impedance models. [DOI: 10.1121/1.1581280]

PACS numbers: 43.20.Rz, 43.20.Tb [LLT]

I. INTRODUCTION

Perforated elements are extensively used in mufflers for the intake and exhaust systems of various fluid machines. Except for some cross-flow elements, which are occasionally used for reducing the propagated noise in spite of the large increase in back pressure, perforated elements are usually exposed to grazing medium flow and it is known that the involved flow, usually of low speed, greatly influences the acoustic impedance of the perforates. The grazing flow in a muffler is usually a fully developed turbulent one and the mean flow velocity is of concern in taking account of the flow effects. Accurate knowledge of the acoustic impedance of a perforate is essential in the prediction of the acoustic performance of the whole perforated element in a muffler system.

The acoustic impedance of a perforate without flow is determined by the relation between the acoustic pressure and the particle velocity inside a perforate, which is corrected by the effect of added mass in the opening. Rayleigh[1] suggested a model whereby the air particles vibrate in unison as an incompressible air piston excited by the incident wave, when the thickness of the perforate is less than a wavelength of an incident wave. In defining the kinetic energy of the air piston, the concept of the end corrected length was introduced. Crandall[2] expressed the orifice impedance by the wave number and wavelength of the viscous Stokes wave, which takes account of the thermal conduction at the inner wall of perforates. Sivian[3] derived the reactance formula from the assumption of Rayleigh, of which the air particles in an orifice act as a group like a rigid air piston. It was suggested that the acoustic resistance of an orifice comes from two viscous loss mechanisms: One is the internal viscous loss of the orifice and the other is the external one. In the external loss part, the viscosity was replaced by a value appropriate to a nonthermally conducting medium, in which the end-corrected orifice length was used. Ingard[4] derived a reactance formula by using the thickness of a viscous boundary layer. Sivian[3] refined the specific resistance formula by classifying the physical conditions into Poiseuille and Helmholtz regimes. The energy loss from the thermal conduction was assumed to be negligible compared with that from the viscous force and the resistance due to viscous losses was expressed by using a surface resistance. Melling[5] included the effect of mutual interaction between neighboring orifices in the theoretical model. The effect of orifice interaction on impedance was given as a function of orifice spacing.[6] Sullivan and Crocker[7] measured the impedance of perforated elements in a concentric tube resonator and then modified the theoretical model of Melling[5] for dealing with various diameters and thicknesses of practical orifices. The impedance was classified into linear and nonlinear regimes by using the concept of a critical velocity relationship with frequency. All of the aforementioned impedance models were restricted to the condition of no mean flow of the background medium.

In the presence of flow, rigorous mathematical modeling of the mechanisms which determine the acoustic impedance

[a)]Author to whom correspondence should be addressed. Electronic mail: ihih@sorak.kaist.ac.kr

of a perforate is extremely difficult. Consequently, most of the existing impedance models considering the flow effect are empirical and, also, most of them are only for the grazing mean flow condition. Garrison[8] suggested an empirical impedance model in the presence of grazing mean flow, in which the impedance formula requires the tangential Mach number of the grazing flow over the orifice. Ronneberger[9] measured the acoustic impedance of perforates under the grazing flow condition varying the Strouhal number. However, the parametric study was incomplete due to the use of insufficient perforated elements. Dean[10] used the two-microphone technique in the measurement of orifice impedance and investigated the characteristics of factors that influence the measurement precision. Varying the mean flow and perforate conditions, the acoustic impedance of a perforated liner having the honeycomb in the backing cavity was measured. Bauer[11] considered both grazing and cross-flow for a perforated liner with a backing cavity and suggested an empirical model. No flow effect was considered in the reactance, but the porosity of the orifices was included in the resistance and reactance, thus taking account of the effect of orifice interactions. Hersh *et al.*[12] studied experimentally multiple orifices under the influence of grazing flow. They found that the reactance is quite dependent on the orifice spacing, particularly when the orifices are aligned in parallel to the grazing flow direction. In the study on the semi-empirical fluid mechanical model of a Helmholtz resonator, it was noted that the incident sound field is connected with the sound field of the backing cavity by an orifice discharge coefficient. Sullivan[13] modified the impedance model without flow into a new one considering the through-flow velocity in the resistance, referring to Garrison's impedance model.[8] Jayaraman and Yam[14] modified the no-flow impedance model of Sullivan and Crocker[7] using the grazing flow Mach number. The grazing flow effect was considered only in the resistance model. Goldman and Panton[15] and, later, Goldman and Chung[16] studied the influence of the boundary layer on the wall of a flow duct upon the acoustic impedance of an orifice. It was shown that the orifice impedance could be expressed in terms of the friction velocity, which is a measure of properties in the boundary layer. Rao and Munjal[17] suggested an empirical model of the acoustic impedance of an orifice. The acoustic impedance was measured by the impedance tube technique for a variety of perforated tubes while varying the involved geometrical and flow parameters. The range of parameters was based on the practical data span employed in typical automotive exhaust mufflers. However, the pressure near perforates in the flow duct was assumed to be the same as the pressure in the flow duct without perforates and flow effects were not considered in the reactance model. Kooi and Sarin[18] investigated experimentally to find the impedance variation of resonator arrays for different boundary layer velocity profiles under grazing flow conditions. The orifice impedance was expressed as a function of the inverse Strouhal number associated with the orifice diameter or thickness. Cummings[19] measured the acoustic impedance of some types of perforates and expressed the results in terms of the flow resistance and the end-correction ratio. Kirby and Cummings[20] measured the impedance of perforates with and without porous backing in the cavity.

Although there are many analytical and empirical orifice impedance models, as mentioned above, the predicted acoustical performance adopting any one of them sometimes shows a large discrepancy with the measured result in some cases. In this study, in order to obtain a new practical and precise empirical impedance model under grazing flow conditions, the acoustic impedance of circular orifices is measured with a variation of the involved parameters under very carefully tested and controlled measurement conditions. The parameters involved in determining the acoustic impedance of an orifice are comprised of the orifice diameter, orifice thickness, perforation ratio, mean flow velocity, and frequency. The range of involved parameters is selected to cover the practical data span of perforates in the typical exhaust systems of internal combustion engines, thus extending the applicable parameter range compared to the previous impedance models. The empirical model is obtained by using nonlinear regression analysis of the various results of the parametric tests. The proposed empirical model of orifice impedance, with a very high correlation coefficient, is applied to the prediction of the transmission loss of concentric resonators, which have geometric configurations typical of the acoustically short and long through-flow resonators.

II. IMPEDANCE MODEL

The acoustic impedance of perforated elements is defined as the ratio of pressure differences between front and rear of the perforated elements to a particle velocity inside of the perforated elements as

$$Z=\frac{p_f-p_r}{u'}, \tag{1}$$

where Z is the acoustic impedance, p_f is the sound pressure in front of the perforates, p_r is the sound pressure in the rear of the perforates, and u' the averaged particle velocity in the perforates.

A. Review of some current impedance formulas in the presence of flow

Some of the aforementioned impedance models have been employed in the prediction of perforated resonator elements in practical muffler systems. Garrison[8] proposed the following empirical model considering effects of grazing flow:

$$Z=(1+1.9\mathrm{M})R_0+j(1-1.65\mathrm{M})I_0, \tag{2}$$

where R_0 and I_0 are the resistance and reactance, respectively, in the absence of mean flow. M is the Mach number of the grazing mean flow.

Bauer[11] suggested an empirical impedance model containing the influence of the cross flow velocity as follows:

$$Z_n=\left[\left(\frac{\sqrt{8\mu\rho_0\omega}}{\rho_0 c_0\sigma}\right)\left(1+\frac{t}{d}\right)+\frac{0.3\mathrm{M}}{\sigma}+\frac{1.15(\overline{U_0}/c_0)}{\sigma}\right]+j\left[\frac{k(t+0.25d)}{\sigma}\right]. \tag{3}$$

Here, $\rho_0 c_0$ is the characteristic impedance of the medium, in which ρ_0 and c_0 are the density and the speed of sound, respectively. Z_n is the normalized acoustic impedance which is divided by $\rho_0 c_0$, $\overline{U_0}$ is the cross flow velocity through the perforate, μ is the viscosity coefficient of the medium, σ is the porosity of the perforate, d is the orifice diameter, t is the thickness of the orifice, ω is the circular frequency, and k is the wave number.

Sullivan[13] modified his impedance model for the no flow condition[7] to include the effect of flow by referring to the empirical model of Garrison.[8] Only the cross flow velocity was used for the resistance model, while the reactance, which decreases with the increasing flow velocity, was assumed to be the same as that for the no-flow condition:

$$Z_n = [2.57\overline{U_0}/c_0 + jk(t+0.75d)]/\sigma. \tag{4}$$

Jayaraman and Yam[14] modified the impedance model of Sullivan[7] in which the grazing flow velocity was influential only upon the resistance model:

$$Z_n = (0.514d\mathrm{M}/t\sigma + j4.8\times 10^{-5} f)/\sigma. \tag{5}$$

Here, f is the frequency.

Rao and Munjal[17] determined the acoustic impedance using the impedance tube technique and proposed the following empirical model, considering all involved parameters such as diameter and thickness of the perforate, porosity, and grazing mean flow velocity. Flow effect was considered only in the resistance model. The normalized acoustic impedance of an orifice in the perforated pipe was empirically given by

$$Z_n = [7.337\times 10^{-3}(1+72.23\mathrm{M})]/\sigma + j[2.2245\times 10^{-5}\times f(1+51t)(1+204d)]/\sigma. \tag{6}$$

The valid input data ranges were such that $0.05 \leqslant \mathrm{M} \leqslant 0.2$, $0.03 \leqslant \sigma \leqslant 0.1$, $1 \leqslant t \leqslant 3$ mm, $1.75 \leqslant d \leqslant 7$ mm.

Kooi and Sarin[18] proposed the orifice acoustic impedance defined empirically by using the flow resistance and the end correction ratio. The flow resistance denotes the change of resistance induced by flow, which actually means the difference between the measured resistance and the viscous losses in the orifice. The end correction ratio implies the ratio of end correction length at the grazing flow condition to that in the absence of the grazing flow. The empirical equation was given by

$$R_f c_0/fd = (5-t/d)(9.9u_*/fd - 3.2)/4, \tag{7}$$

$$\delta/\delta_0 = 0.92 - 0.75u_*/ft + 0.11(u_*/ft)^2 \quad (0.2 \leqslant u_*/ft \leqslant 3.5), \tag{8a}$$

$$\delta/\delta_0 = -0.04 \quad (u_*/ft > 3.5). \tag{8b}$$

Here, R_f denotes the area-averaged flow resistance of a single orifice, which can be obtained by multiplying the porosity with the flow-induced resistance change of the whole perforated specimen, δ is the end correction length in the presence of grazing mean flow, and δ_0 is the end correction length in the absence of grazing flow.

Cummings[19] modified the foregoing empirical impedance model and suggested the following model:

$$R_f c_0/fd = (12.52(t/d)^{-0.32} - 2.44)(u_*/fd) - 3.2, \tag{9}$$

$$\delta/\delta_0 = 1 \quad (u_*/ft \leqslant 0.12d/t), \tag{10a}$$

$$\delta/\delta_0 = (1+0.6t/d)\exp[-(u_*/ft - 0.12d/t)/(0.25+t/d)] - 0.6t/d \quad (u_*/ft > 0.12d/t). \tag{10b}$$

Later, Kirby and Cummings[20] measured the acoustic impedance of perforates with and without porous backing. An empirical impedance model of perforates without porous backing was given by

$$R_f c_0/fd = (26.16(t/d)^{-0.169} - 20)(u_*/fd) - 4.055, \tag{11}$$

$$\delta/\delta_0 = 1 \quad (u_*/ft \leqslant 0.18d/t), \tag{12a}$$

$$\delta/\delta_0 = (1+0.6t/d)\exp[-(u_*/ft - 0.18d/t)/(1.8+t/d)] - 0.6t/d \quad (u_*/ft > 0.18d/t). \tag{12b}$$

In the aforementioned models considering turbulent effects, the normalized acoustic resistance R_n is composed of the viscous loss effect R_0 in the orifice and the flow resistance R_f as follows:

$$R_n = R_f + R_0. \tag{13}$$

The viscous loss R_0 is defined as

$$R_0 = \frac{\sqrt{8\nu\omega}}{c_0}\frac{t}{d}, \tag{14}$$

where ν denotes the kinematic viscosity. The normalized acoustic reactance I_n can be obtained from the end correction ratio δ/δ_0 by the following relation under the assumption that the end correction length without flow is approximately equal to $0.85d$:

$$I_n = k\left[t + \frac{\delta}{\delta_0}(0.85d)\right]. \tag{15}$$

The major reason why many previous impedance models are listed and described here is that the parameters involved in various impedance formulas of the earlier models[8,11,13,14,17] definitely differ from those of the later models.[18–20] The main intention is to integrate both types in a unified format so that the resultant impedance model can be used easily in practical situations, in particular for the exhaust silencing system of internal combustion engines.

B. Experimental setup

Figure 1 shows the experimental setup for the measurement of acoustic impedance of a circular orifice. A compressor, through its ancillaries for flow regularization, supplied the airflow and the flow velocity was measured by the Pitot tube and the micro-manometer. The mid-stream flow Mach number was ranged from 0 to 0.2, which corresponds to the typical range of flow velocity in the intake/exhaust systems of internal combustion engines of cars. The source sound was generated by a 4-in., 50-W loudspeaker mounted on an upstream side-wall, with a lower cutoff frequency of 180 Hz. A square acrylic duct, 30×30 mm^2 in internal size and 5 mm in thickness, was used for easy installation of the perforated

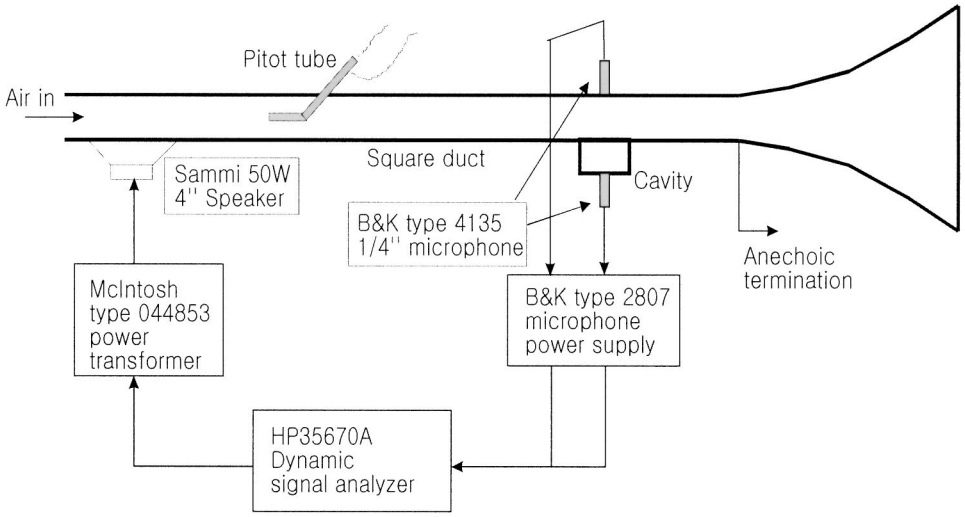


FIG. 1. Experimental setup for the measurement of acoustic impedance of perforates.

panel specimen and the backing cavity, of which the sectional dimension was 30×30 mm^2 with varying length to adapt for the specimen length. Quarter-inch microphones (B&K 4135) were flush mounted on the bottom of the cavity and in front of the perforated specimen plate in the square main duct. The frequency response function between the two microphones was measured and the phase matching was tested before the main experiments. The phase matching of the two sets of microphones and amplifiers was within 0.5° over the frequency range of the measurements and the remaining small mismatched phase was compensated in the post processing. Generation of source signal and analysis of measured signal were all performed by using a multi-channel signal analyzer (HP 35670A). The background noise level was less than 48 dB for all frequencies during the experiments. Figure 2 shows a detailed view of the setup of a perforated specimen, backing cavity, and measurement sensors. The impedance measurement was based on the two-microphone technique used in the previous works.[10,18–20] It was assumed that a planar wave was propagating in the cavity and that the pressure measured at the square duct was equal to the pressure under the perforated plate. The impedance was obtained by the frequency response function between microphones using the following relation: across the perforates, the transfer matrix can be written as

$$\begin{Bmatrix} p_D \\ u'_D \end{Bmatrix} = \begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} p_{C0} \\ u'_{C0} \end{Bmatrix} \equiv [T_{D0}] \begin{Bmatrix} p_{C0} \\ u'_{C0} \end{Bmatrix}, \tag{16}$$

where p is the acoustic pressure, u' means the average particle velocity over the orifice, and the subscripts D and $C0$ refer to the region below and above the perforated plate, respectively. Because only the plane wave propagates in the cavity, the transfer matrix between two points $C0$ (mouth) and $C1$ (bottom) in the cavity, as depicted in Fig. 2, can be expressed as

$$\begin{Bmatrix} p_{C0} \\ u'_{C0} \end{Bmatrix} = \begin{bmatrix} \cos kL_C & j\rho_0 c_0 \sin kL_C \\ j\dfrac{1}{\rho_0 c_0} \sin kL_C & \cos kL_C \end{bmatrix} \begin{Bmatrix} p_{C1} \\ u'_{C1} \end{Bmatrix} \equiv [T_{01}] \begin{Bmatrix} p_{C1} \\ u'_{C1} \end{Bmatrix}, \tag{17}$$

where L_c denotes the depth of the cavity. Finally, one can obtain the normalized impedance of the perforated plate from the frequency response function between the two microphones (1 and 2) as follows:

$$Z_n = Z/\rho_0 c_0 = \frac{[H_{12} - \cos kL_C]}{j \sin kL_C}. \tag{18}$$

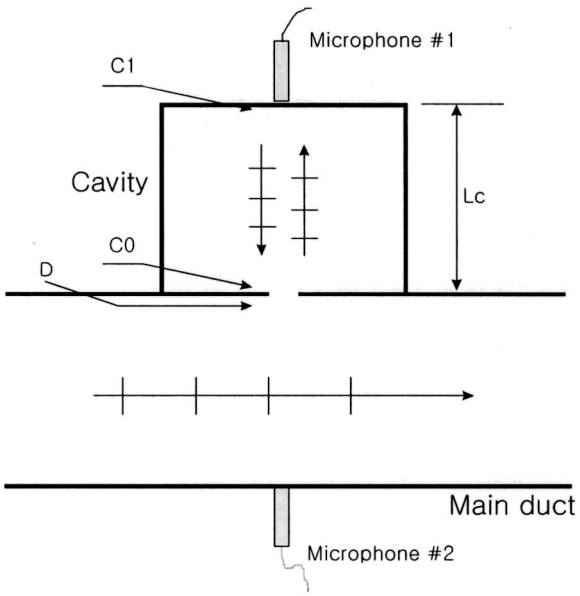


FIG. 2. Detail setup near perforates and cavity.

C. Preliminary measurements

The porosity of a perforated resonator can be defined in the following two ways: one definition is the ratio of the whole perforated area to the whole surface area; another one is the ratio of a single perforated area to the effective area, which is formed by the directional distances from neighboring orifices. For a small number of clustered orifices on the plate or pipe, the latter definition of porosity is valid, in which the total effective area encompasses the region that includes the orifices and the outer extended region given by the half distance between two adjacent orifices. In this situ-

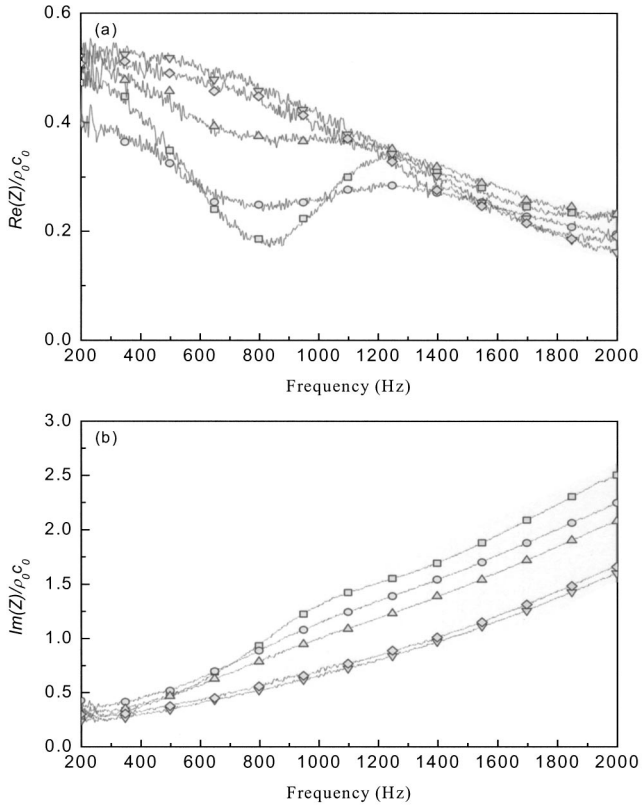


FIG. 3. Impedance of perforates due to the spacing of orifices at M=0.1, σ=7%, d=3 mm, t=1 mm: —□—, ds=4 mm; —○—, ds=4.5 mm; —△—, ds=5 mm; —▽—, ds=7 mm; —◇—, ds=9 mm. (a) Specific acoustic resistance, (b) specific acoustic reactance.

ation, if the distances between two directions are different or inhomogeneous, the latter definition of porosity is different from the former one. In the actual service condition of typical exhaust mufflers, the orifices are nearly uniformly distributed on the resonator inner pipe: in this case, the two definitions are essentially same. In order to clarify and sort out the problem, a preliminary experiment was carried out. The same nine orifices with 3 mm in diameter were made on a 30×30 mm^2 plate, but the distance between two adjacent orifices was varied from 4 to 9 mm in both x and y directions. Figure 3 shows the measured impedance varying the hole-to-hole distance ds (M=0.1, σ=7%, d=3 mm, t=1 mm). It can be seen that the measured resistance value has nearly the same frequency trend and magnitude except for cases with small hole-to-hole separation distances such as 4 and 4.5 mm. This result implies that the energy dissipation in an orifice is also influenced by the neighboring orifices when the orifices are very close to each other. The measured reactance decreases in general with an increase of the separation distance. However, for orifices with very small separation distance such as ds=4 mm, the measured reactance is smaller than expected from the general trend of reactance for frequencies lower than about 600–700 Hz. This means that the reactance is also influenced by the presence of other neighboring orifices when they are located very close to each other, in particular when the frequency is low. From this experimental result, one may conclude that the porosity of a orifice can be defined by the ratio of the whole orifice area to the whole surface area without considering the mutual interactions between orifices if the hole-to-hole separation distance is larger than twice the orifice diameter.

There might be cases for which the perforated region is quite long that would cause a doubt about the basic measurement concept underlying Eq. (18). It is a premise of Eq. (18) that the pressure on each side of the perforate is the same at every orifice, since the pressure is regarded as an average one. The actual measurement is done with a cluster of similar orifices in order to minimize the experimental error. For this reason, one tends to test as many orifices as possible for a given specimen, but the problem in measuring the impedance for a cluster of similar orifices is that a rather long array of orifices will be distributed in the axial direction of the main duct. This will cause a problem in defining the transfer function as in Eq. (18) between two microphones, of which the situation is illustrated in Fig. 2. As mentioned before, the total size of the perforated plate was 30×30 mm^2 in general, but, for the purpose of testing the basic premise in the measurement of transfer functions, samples of size 30×80 mm^2, which have the same geometrical parameters (i.e., porosity, diameter, thickness) with small-size sample plates, were also measured. Equation (18) contains the frequency response function between two microphones located at the main duct and the backing cavity wall in the average sense, by assuming that each measured pressure represents the pressure of one side of an orifice. When the orifices on 30×80 mm^2 plate specimens are considered, the aforementioned average pressure concept can no longer be valid, because the orifices in the plate are axially distributed to an extent whereby they cannot be regarded as clustered within a small region. For this reason, in the case of a 30×80 mm^2 sample plate, three sets of frequency response functions ($H_{aa'}$,$H_{bb'}$,$H_{cc'}$) were measured from the measurement setup as shown in Fig. 4, in which three microphones are located in the middle and at the two extreme positions in the perforated region, and they were averaged for obtaining the acoustic impedance. It is noted in Fig. 5 that the measured acoustic impedances of two cases, which are different in sample size, are nearly the same. This means that, with proper treatment of transfer functions and by arranging the orifices in a symmetrical manner, one can obtain the acoustic impedance of a single orifice precisely, from basic measurements on a cluster of identical orifices.

If the experiments are accomplished with high sound power, there are nonlinear effects on the acoustic impedance of perforates. Ingard and Ising[21] investigated the acoustic nonlinearity of an orifice. They reported that at sound-pressure levels up to 130 dB, nonlinear effects are occurred in acoustic impedance. Sullivan[13] established acoustic energy dissipation increases owing to the increase in resistance of the perforation with an increase in sound pressure up to 140 dB. To consider the influences of sound pressure level (SPL) on impedance, the impedance of specific perforates were measured with various SPL. The background noise level due to the mean flow supplied by compressor is almost 130 dB. Figure 6 shows effects of an overall sound pressure level of microphone 2, which is established on a main duct. With

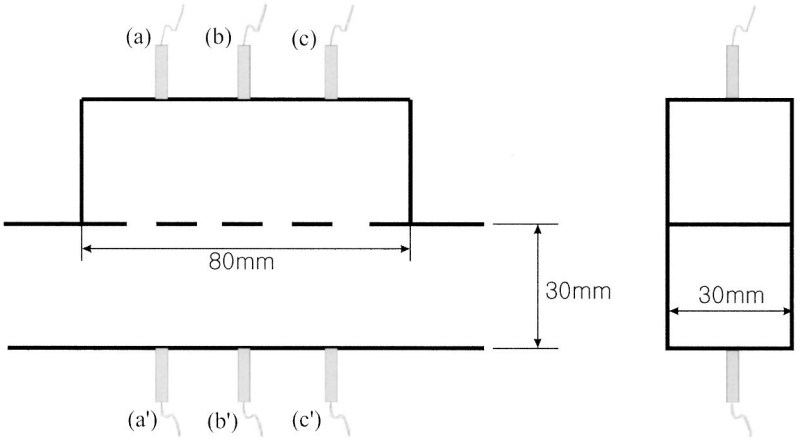


FIG. 4. Experimental sets for checking the effect of size of perforates and cavity.

increasing SPL, resistance tends to increase and reactance decreases a little. This means high SPL makes the impedance of perforates have nonlinear values. The influences of SPL show similar aspects of a Mach number of the grazing flow. The measured impedance data for proposing empirical model with Mach number of 0.1 are got with almost 130 dB. The proposed empirical impedance model is obtained by avoiding nonlinear effects of overall SPL.

D. Measurement results

Parameters that influence the acoustic impedance are porosity, diameter, thickness of an orifice, and the grazing flow velocity. In order to investigate the effect of each parameter, the impedance was measured when varying one parameter of concern while fixing the values of other parameters. Figure 7 shows the measured impedance when the Mach number of grazing mean flow velocity varies from 0 to 0.2 (σ=2.79%, d=4 mm, t=1 mm). It is found that the magnitude of the resistance increases in general as the mean flow Mach number increases, but the rate of decrease of resistance with frequency is nearly the same for every Mach number condition. However, the decreasing trend of resistance with the increase of frequency changes to an increasing one above a certain frequency, at which the resistance is nearly zero. Here, this special frequency, f_0, is defined as the critical frequency in

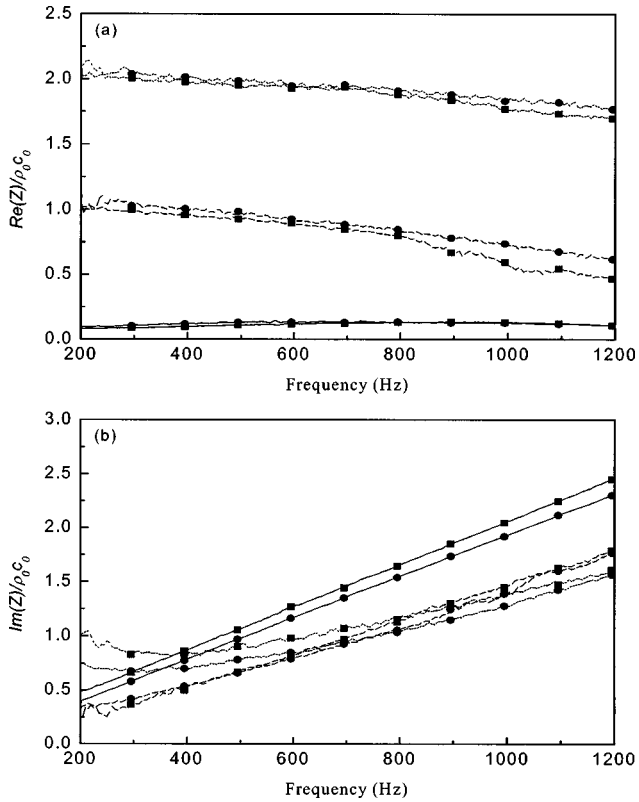


FIG. 5. Impedance of perforates due to the size at σ=4.19%, d=4 mm, t=1 mm: —■—, M=0; —●—, M=0.1; —▲—, M=0.2; black lines, 30×30 mm; gray lines, averaged value using three frequency response functions in case of 30×80 mm. (a) Specific acoustic resistance, (b) specific acoustic reactance.

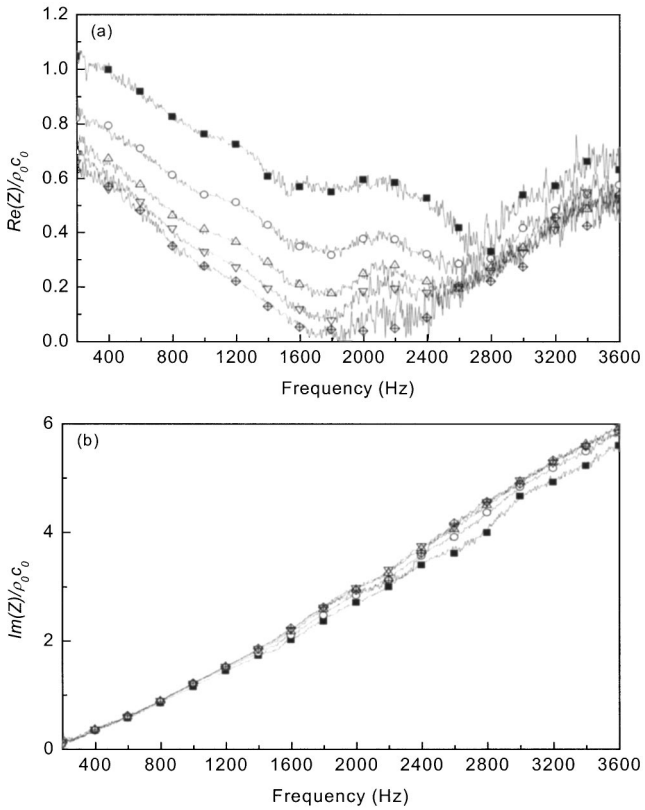


FIG. 6. Measured orifice impedance varying the overall sound pressure level of microphone 2 (M=0.1, d=4 mm, t=1 mm, σ=5.59%): —■—, 145 dB; —○—, 140 dB; —△—, 136 dB; —▽—, 133 dB; —◈—, 130 dB. (a) Specific acoustic resistance, (b) specific acoustic reactance.

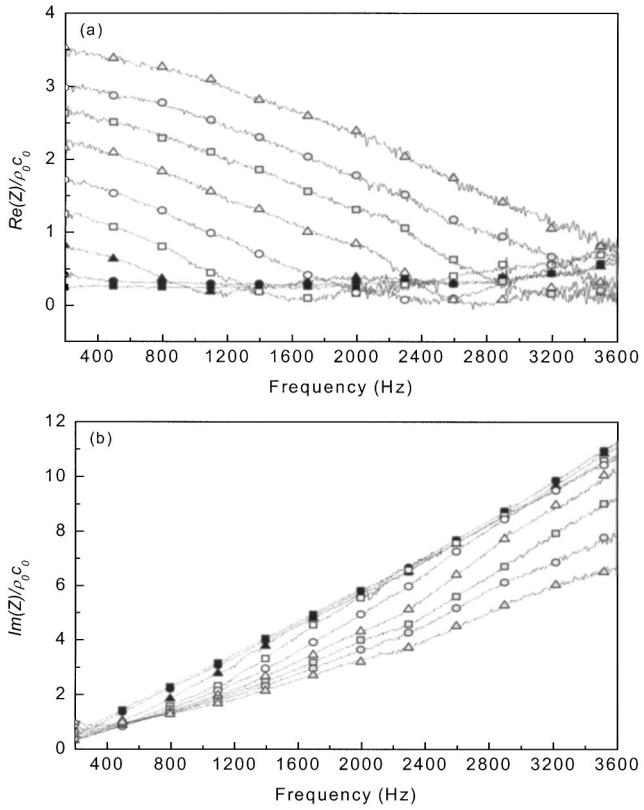


FIG. 7. Measured orifice impedance varying the flow velocity (σ=2.79%, d=4 mm, t=1 mm): —■—, M=0; —●—, M=0.025; —▲—, M=0.05; —□—, M=0.075; —○—, M=0.1; —△—, M=0.125; —■—, M=0.15; —●—, M=0.175; —▲—, M=0.2. (a) Specific acoustic resistance, (b) specific acoustic reactance.

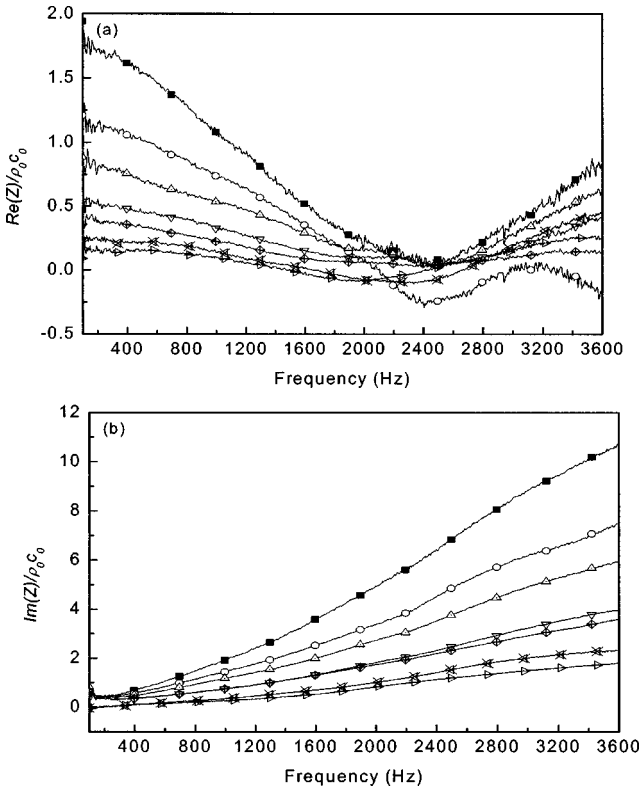


FIG. 9. Measured orifice impedance varying the porosity (M=0.1, d=4 mm, t=1 mm): —■—, σ=2.79%; —○—, σ=4.19%; —△—, σ=5.59%; —▽—, **σ=8.38%**; —◈—, σ=9.79%; —⋊—, σ=15.4%; —▷—, σ=22.3%. (a) Specific acoustic resistance, (b) specific acoustic reactance.

resistance and is different for each flow velocity condition. Reactance increases with frequency and its magnitude generally decreases as the flow Mach number increases, although the measured reactance values are nearly the same at high frequencies when the Mach number is smaller than about 0.1. It is worth mentioning that the critical frequency in resistance can be seen more clearly if one redraw the resistance curves as a function of Strouhal number, $R\omega/U$, as can be seen in Fig. 8. The measured impedance when varying the perforate porosity (M=0.1, d=4 mm, t=1 mm) is shown in Fig. 9, in which the range of porosity is from 2.79% to 22.3%. The spectral values of the resistance show V-shaped curves for all porosities centered at around 2.4 kHz, at which the resistance is nearly the same regardless of the porosity and is very small, near to zero. Also, one can see that, with increasing porosity, both resistance and reactance decrease in general. Figure 10 illustrates the measured impedance when varying the orifice diameter (M=0.1, σ=5.59%, t=1 mm). As the orifice diameter increases, the reactance increases slightly. In a similar manner to Fig. 9, the resistance decreases in general with an increase of orifice diameter, but at low frequencies the resistance decreases with increasing frequency and then increases after passing a certain frequency. Figure 11 shows the effect of orifice thickness (M=0.1, σ=5.59%, d=4 mm). In general, resistance decreases and reactance increases with an increase of the orifice thickness. The overall trend of the resistance is similar to that of Fig. 9, as the curves form V-shapes with a minimum at about 2.4 kHz.

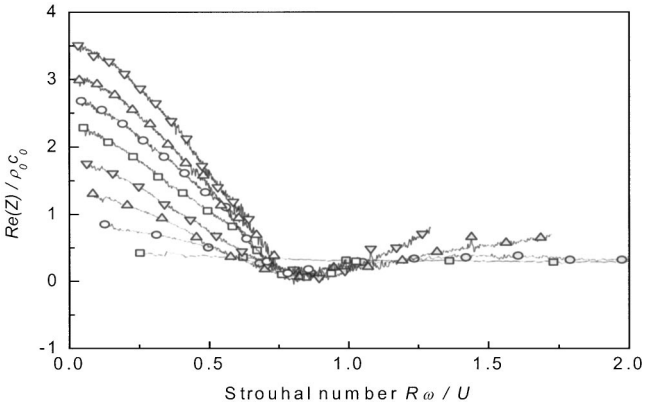


FIG. 8. Measured orifice resistance with respect to Strouhal number varying the flow velocity (σ=2.79%, d=4 mm, t=1 mm): —■—, M=0.025; —●—, M=0.05; —▲—, M=0.075; —▽—, M=0.1; —□—, M=0.125; —○—, M=0.15; —△—, M=0.175; —▽—, M=0.2.

E. Regression curves and empirical impedance model

It was intended to derive an empirical impedance formula from the experimental data of the parametric study. To analyze the experimental data, a commercial statistical calculation package SAS (Statistical Analysis System) was em-

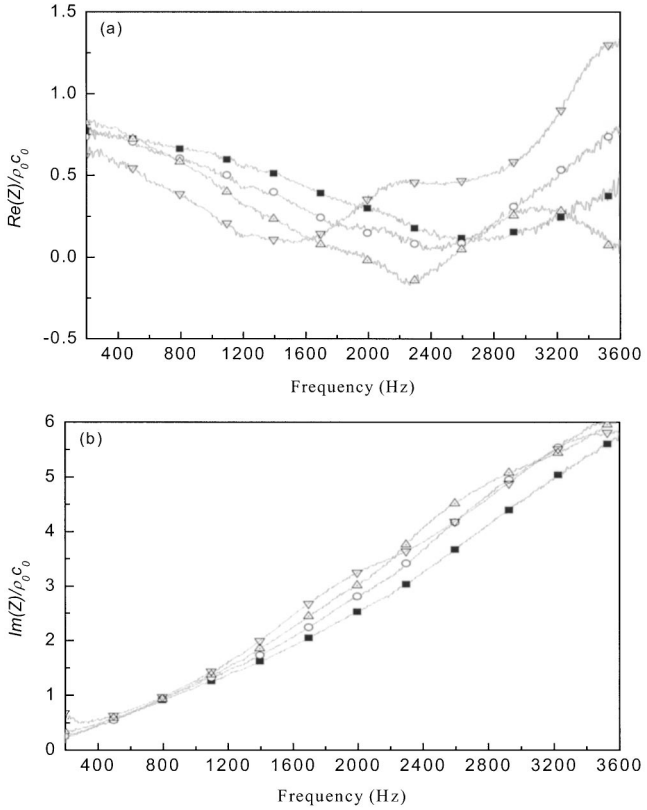


FIG. 10. Measured orifice impedance varying the orifice diameter (M=0.1, σ=5.59%, t=1 mm): —■—, d=8 mm; —○—, d=4.6 mm; —△—, d=4 mm; —▽—, d=3 mm. (a) Specific acoustic resistance, (b) specific acoustic reactance.

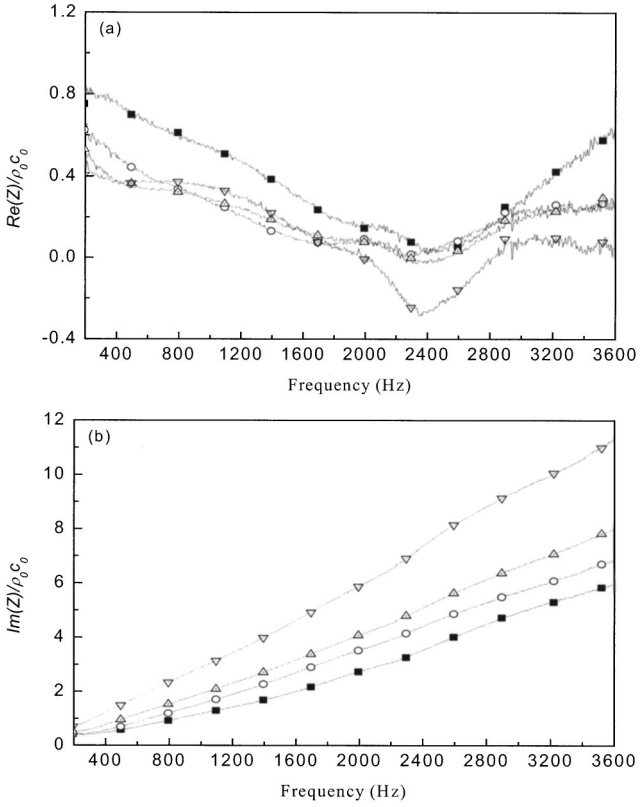


FIG. 11. Measured orifice impedance varying the orifice thickness (M=0.1, σ=5.59%, d=4 mm): —■—, t=1 mm; —○—, t=2 mm; —△—, t=3 mm; —▽—, t=5 mm. (a) Specific acoustic resistance, (b) specific acoustic reactance.

ployed. Because the relations among the involved parameters and the target values, i.e., resistance and reactance, are nonlinear, the nonlinear regression technique was used for deriving the empirical impedance model. In the regression of impedance, the valid data ranges were as follows: $60 \leqslant f \leqslant 4000$ Hz, $0 \leqslant \mathrm{M} \leqslant 0.2$, $2 \leqslant d \leqslant 9$ mm, $1 \leqslant t \leqslant 5$ mm, and $2.79\% \leqslant \sigma \leqslant 22.3\%$. These ranges of parameters encompass the normal conditions of most perforated elements as found in automotive mufflers in practical service.

When the resistance is concerned, the first thing that should be taken into account in the regression model is the critical frequency in resistance, f_0. Experimental results in the preceding section reveal that the critical frequency in resistance depends only on the mean flow velocity and orifice diameter. Consequently, the regression model of f_0 can be given by

$$f_0 = \phi_1 \frac{1+\phi_2 \mathrm{M}}{1+\phi_3 d}, \tag{19}$$

$$\phi_1 = 412, \tag{20a}$$

$$\phi_2 = 104, \tag{20b}$$

$$\phi_3 = 274. \tag{20c}$$

The variables (ϕ_1, ϕ_2, ϕ_3) were obtained from the nonlinear regression of the data containing the effects of f_0, M, and d. The correlation factor R^2 between the regression model and the experimental data is 0.9809. The valid range is such that $0.025 \leqslant \mathrm{M} \leqslant 0.2$, $2 \leqslant d \leqslant 9$ mm.

As mentioned before, the frequency dependence of the resistance is that the resistance decreases with an increase of frequency in the frequency range below f_0, while it increases above f_0. Taking account of the effects of all the other parameters together, an empirical resistance model can be obtained by the regression analysis. The target function is given by

$$R_n = a_0(1+a_1|f-f_0|)(1+a_2\mathrm{M})(1+a_3 d)(1+a_4 t)/\sigma, \tag{21}$$

where R_n denotes the normalized resistance and a_i (i = 0,1,2,3,4) are the regression coefficients. From the nonlinear regression with measured data, the regression coefficients are obtained as

$$a_0 = 3.94e-4, \tag{22a}$$

$$a_1 = 7.84e-3, \tag{22b}$$

$$a_2 = 14.9, \tag{22c}$$

$$a_3 = 296, \tag{22d}$$

$$a_4 = -127, \tag{22e}$$

for which the correlation coefficient R^2 is 0.8458. Figures 12(a)–15(a) depict the regressed resistance curves from the measured data in Figs. 7(a) and 9(a)–11(a). It is thought that the negative value of measured resistance in Figs. 7 and

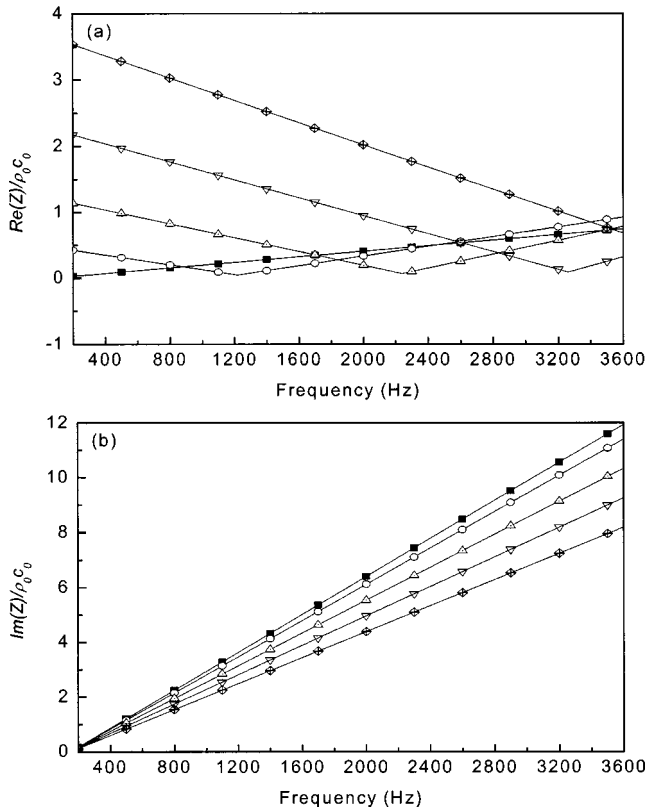


FIG. 12. Orifice impedance model varying the flow velocity (σ=2.79%, d =4 mm, t=1 mm): —■—, M=0; —○—, M=0.05; —△—, M=0.1; —▽—, M=0.15; —◈—, M=0.2. (a) Specific acoustic resistance, (b) specific acoustic reactance.

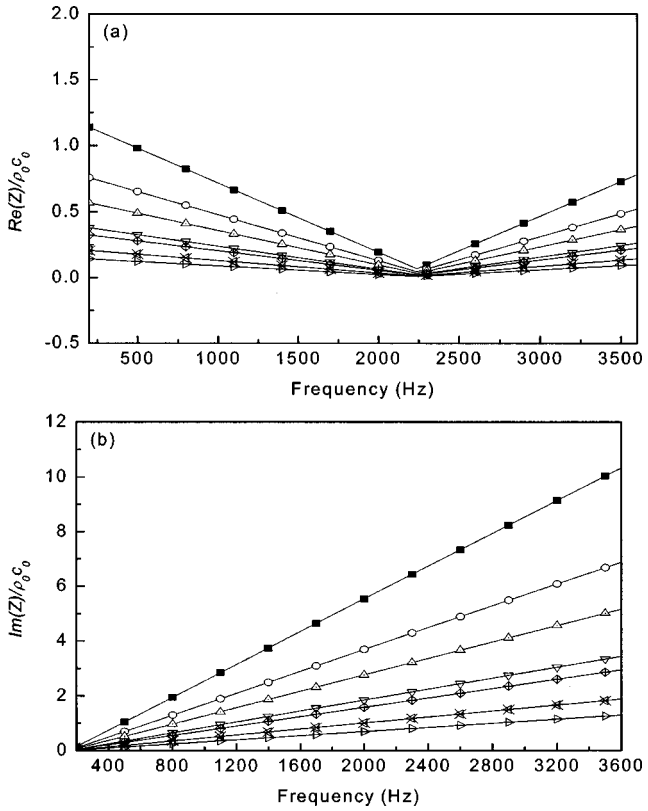


FIG. 13. Orifice impedance model varying the porosity (M=0.1, d =4 mm, t=1 mm: —■—, σ=2.79%; —○—, σ=4.19%; —△—, σ=5.59%; —▽—, σ=8.38%; —◈—, σ=9.79%; —⋈—, σ=15.4%; —▷—, σ=22.3%. (a) Specific acoustic resistance, (b) specific acoustic reactance.

9–11 is generated from the self-noise of perforates. This means that the condition for having negative resistance is quite limited to some specific parametric values for satisfying the onset condition of aerodynamic sound generation. In accordance to this argument, most of the measured impedance data near the critical frequency are positive although they are very near to zero. Because the resistance is negative for very limited cases and also the negative deviation is small, negative cases are not taken into account in deriving the empirical model, which should be as simple as possible to be used for practical engineering applications.

When considering the reactance, the effects of Mach number and frequency are considered first. Measured impedance data reveals that the reactance increases as the frequency increases and generally decreases as the flow Mach number increases. It is found that the measured reactance values are nearly the same at high frequencies when the flow Mach number is smaller than about 0.1. Such a complicated behavior cannot be described by a simple first-order polynomial as a function of frequency and Mach number. For this reason, the following empirical model of normalized reactance is proposed, taking account of the effects of all parameters together:

$$I_n = b_0(1+b_1 d)(1+b_2 t)(1+b_3 \mathrm{M})(1+b_4 f)/\sigma. \quad (23)$$

Here, b_i $(i=0,1,...,4)$ are regression coefficients, which are determined by the nonlinear regression analysis with measured data as

$$b_0 = -6.00e-3, \quad (24a)$$

$$b_1 = 194, \quad (24b)$$

$$b_2 = 432, \quad (24c)$$

$$b_3 = -1.72, \quad (24d)$$

$$b_4 = -6.62e-3. \quad (24e)$$

In this case, the correlation coefficient R^2 is 0.9711. Figures 12(b)–15(b) show the regressed reactance curves from the measured data in Figs. 7(b) and 9(b)–11(b).

It is noted that some differences can be observed between predicted impedance values and measured data. Intensive efforts were directed to fit the data appropriately, but the results showed that the problem could not be solved at the moment. It is concluded that, basically speaking, making a unique regression model from a variety of data is very much prone to accompanying errors because a number of parameters are involved in obtaining the regression model. In addition, measured resistance data show negative values in a few cases (which are physically meaningful) though the most part of measured resistance has no negative value. It could be a cause for errors.

In general, the relation between the impedance of perforated plate and the hole impedance is such that

$$Z_p = \frac{Z_s}{\sigma}, \quad (25)$$

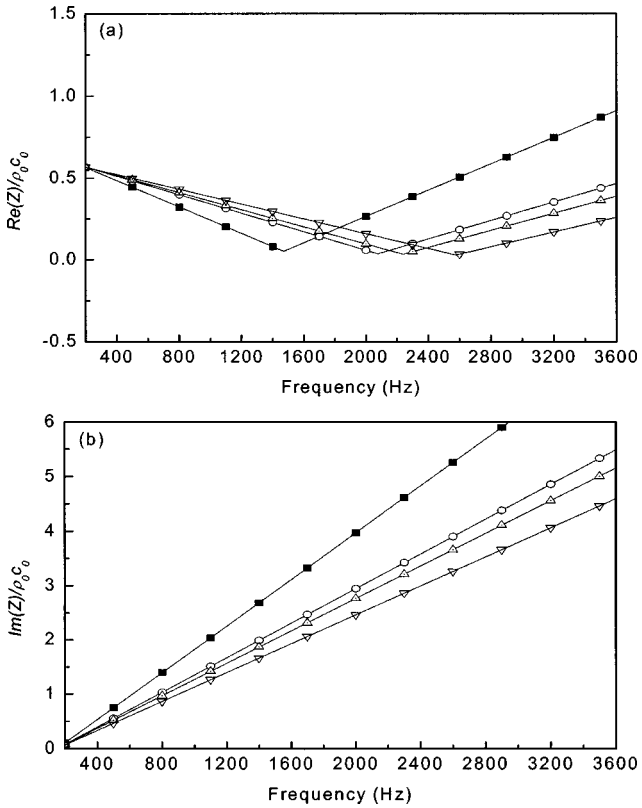


FIG. 14. Orifice impedance model varying the orifice diameter (M=0.1, σ=5.6%, t=1 mm: —■—, d=8 mm; —○—, d=4.6 mm; —△—, d=4 mm; —▽—, d=3 mm. (a) Specific acoustic resistance, (b) specific acoustic reactance.

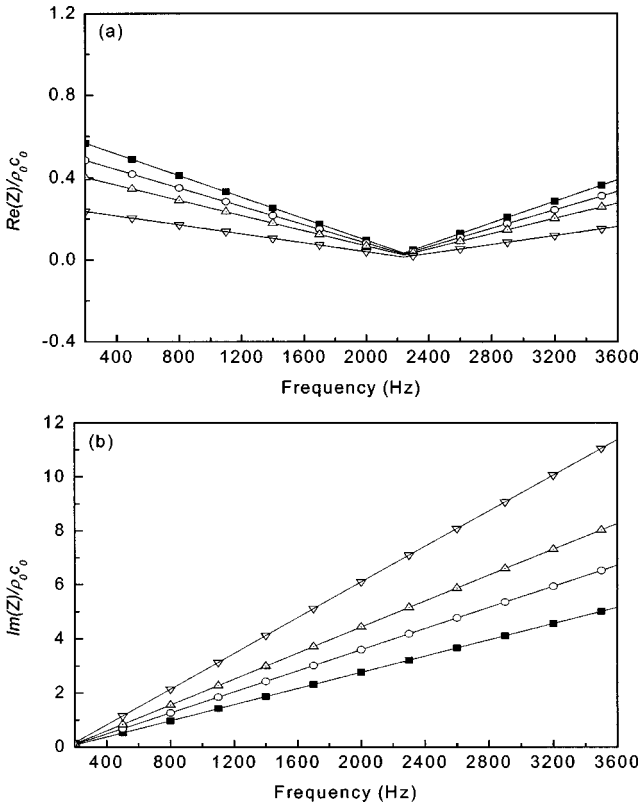


FIG. 15. Orifice impedance model varying the orifice thickness (M=0.1, σ=5.6%, d=4 mm: —■—, t=1 mm; —○—, t=2 mm; —△—, t=3 mm; —▽—, t=5 mm. (a) Specific acoustic resistance, (b) specific acoustic reactance.

where Z_p is the impedance of perforated plate and Z_s is the hole impedance. In Fig. 11, all parameters except porosity are same, which means that the hole impedance is identical for all cases. Therefore, when we derive the hole impedance from the measurement, the porosity does not matter as long as the added mass regions of the adjacent holes do not overlap or such an overlapping does not have any significant effect to the result. Such a case is limited to very high porosity specimens, which is quite hard in finding practical applications from the reason of the structural vulnerability and manufacturing difficulty of perforated plate or tube. From the aforementioned relation in Eq. (25), one can say that the impedance of perforated plate decreases if the porosity increases. Also, it should be mentioned here that experiments and proposed impedance model have never been concentrated on the low porosity conditions only. The porosity of 15% is not a low porosity one. For silencer designers, it has been well known that the acoustic performances of perforated resonators are nearly the same as those without perforated parts, i.e., just simple expansion chambers, when the porosity is larger than 20%, in particular for the through-flow type resonator. Further to this information, the field experience of many practitioners tells that the acoustic performance is very similar for every perforated resonator if their porosity is larger than about 15%–16%, which also shows similar acoustic performance with corresponding simple expansion chamber. The reason why the perforated resonator having high porosity larger than 15%–20% is as follows: first, the perforated pipe can be used as a flow guide to reduce the flow-generated noise and back pressure; second, the perforated pipe can be used for strengthening the muffler structure by connecting inlet and outlet; third, the perforated pipe can be used for protecting and forming the liner material inside the annular cavity for reactive-resistive mufflers; and, fourth, the high-porosity section allows for the airflow easily resulting in a low back pressure through the element, in particular for the cross-flow-type resonators.

F. Comparison of present model with previous impedance models

Figures 16 and 17 compare the proposed impedance model with previous ones for two different mean flow velocity conditions of M=0.1 and 0.2 (σ=5.59%, d=4 mm, t=1 mm). In expressing the impedance model of Refs. 18–20, the following relations between resistance and the friction velocity are used:

$$u_* = \sqrt{\tau_w/\rho} \tag{26}$$

$$\tau_w = \rho U^2 \lambda/8, \tag{27}$$

$$\lambda = 0.306\,\mathrm{Re}^{-1/4} \quad (10^4 < \mathrm{Re} < 10^5), \tag{28}$$

$$u_* = \sqrt{U^2(0.306)\mathrm{Re}^{-1/4}/8}, \tag{29}$$

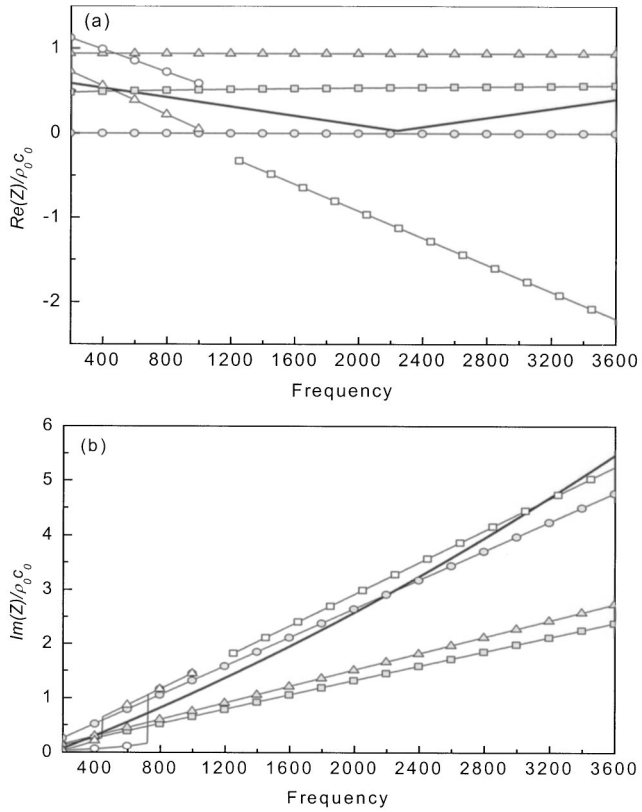


FIG. 16. A comparison of various orifice impedance models at M=0.1, σ=5.59%, d=4 mm, t=1 mm: ———, present model; —■—, Bauer's model; —●—, Sullivan's model; —▲—, Rao and Munjal's model; —□—, Kooi and Sarin's model; —○—, Cummings' model; —△—, Kirby and Cummings' model. (a) Specific acoustic resistance, (b) specific acoustic reactance.

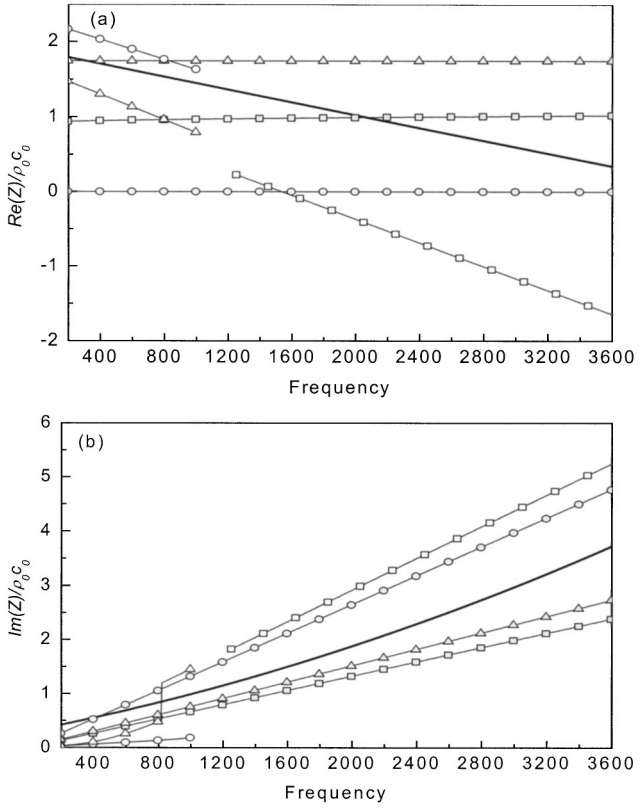


FIG. 17. A comparison of various orifice impedance models at M=0.2, σ=5.59%, d=4 mm, t=1 mm: ———, present model; —■—, Bauer's model; —●—, Sullivan's model; —▲—, Rao and Munjal's model; —□—, Kooi and Sarin's model; —○—, Cummings' model; —△—, Kirby and Cummings' model. (a) Specific acoustic resistance, (b) specific acoustic reactance.

$$\frac{u_*}{fd} = \frac{U}{\pi f d}\pi\sqrt{\frac{(0.306)\mathrm{Re}^{-1/4}}{8}} = \frac{1}{\mathrm{St}}\pi\sqrt{\frac{(0.306)\mathrm{Re}^{-1/4}}{8}} \equiv \frac{C}{\mathrm{St}}. \tag{30}$$

Here, u_* is the friction velocity, τ_w is the wall shear stress, λ is the friction factor, Re is the Reynolds number, St is the Strouhal number, and C is a constant. In Eq. (30), the valid range of Re should be obtained from each recent result in Refs. 18–20. The friction velocity term, u_*/fd, can be regarded as a constant divided by the Strouhal number. Fujita[22] yielded the experimental data on both the perimetral variation of τ_w and the friction factor as a function of Reynolds number based on the hydraulic diameter of main square ducts having either rough or smooth walls. In his friction factor model, the valid range of Reynolds number was from 10^4 to 10^5. From Eq. (30), the friction velocity term, u_*/fd, can be regarded as a constant divided by the Strouhal number. The impedance model of Kooi and Sarin[18] was proposed within the third-octave center frequencies from 1250 to 5000 Hz. The hydraulic diameter of the main duct was 0.3 m and the Reynolds number varied between 3.4×10^6 and 1.0×10^7 in their experiments. The hydraulic diameter of the main duct used for deriving the Cummings' model[19] was 0.0381 m and the Reynolds number range was 3.8×10^4 to 1.8×10^5. The impedance model of Kirby and Cummings[20] was proposed over the friction velocity range of 0.476–2.192, in which the hydraulic diameter of main duct was 0.072 m. Using Eq. (29) with their experimental data, the grazing mean flow velocity U can be calculated as 11–63 m/s and the Reynolds numbers range from 1.8×10^5 to 1.0×10^6. Impedance models of Cummings[19] and Kirby and Cummings[20] were proposed over the frequency range from 70 Hz to 1 kHz. It should be mentioned here that all empirical impedance models are valid within the experimental frequency range of each model. For the case of M=0.1, the resistance in this study has a quite different frequency pattern as compared with others. The reactance curve is very similar to that of the Sullivan's model[13] in Eq. (4), which was assumed to be the same as that of the no-flow condition. For

TABLE I. Specification of measured concentric tube resonators (d_1 = 0.032 m, d_2=0.11 m, t=0.002 m). d_1 and d_2 indicate the diameter of main duct and that of resonator, respectively.

Type	l (m)	σ (%)	d (m)	l/d_2
1	0.2	4.7	0.004	1.81
2	0.2	9.0	0.006	1.81
3	0.2	14.7	0.008	1.81
4	0.04	5	0.004	0.36
5	0.04	10	0.006	0.36
6	0.04	15	0.006	0.36

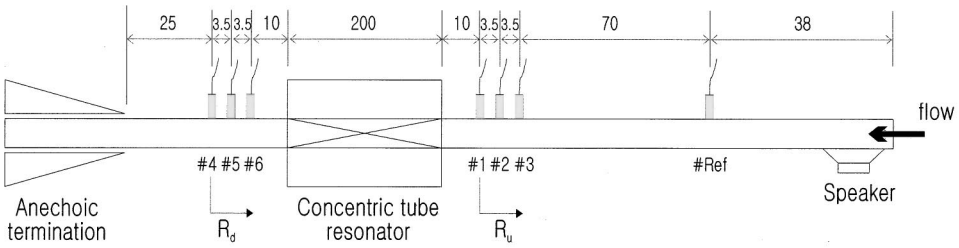


FIG. 18. Experimental setup for measuring the transmission loss of concentric tube resonator.

the case of M=0.2, the increased resistance value shows a monotonically decreasing trend in the frequency range considered. In the very-low-frequency region, the average value is similar to the resistance curve of the Bauer's model[9] in Eq. (3). The reactance is decreased in comparison with the case of M=0.1 and its magnitude is in between the two curves of the Sullivan's model[13] in Eq. (4) and Rao and Munjal's model[17] in Eq. (6). In Figs. 16 and 17, it is noted that the empirical model of Rao and Munjal[17] gives the largest resistance value and the small reactance value. In contrast, the Kooi and Sarin's model[18] yields the smallest resistance value and the largest reactance value among all the impedance models.

III. TEST EXAMPLES: COMPARISON OF TRANSMISSION LOSS

For the purpose of testing the validity and applicability of the derived empirical impedance model, the transmission losses of various concentric tube resonators are predicted by employing the present and previous impedance models and are then compared with the measured ones. The transmission loss can be predicted by solving the coupled second-order ordinary differential equations relating the acoustic fields of an inner tube and an outer cavity through orifices.[14,23–26] Six types of resonators were selected as specified in Table I, of which the length to diameter ratio is either 0.36 (i.e., an acoustically short chamber) or 1.81 (i.e., an acoustically long chamber)[27] and the porosity varies from 5% to 15%.

A. Measurement of transmission loss

The measurement setup for the transmission loss is illustrated in Fig. 18. The diameter of main duct was 0.032 m and the diameter of resonator was 0.110 m. The flow with M=0–0.2 was supplied by an air compressor through a flow regulating section; the mean flow velocity was measured by a Pitot-tube and a micro-manometer. To eliminate the negative effect of strong reflective waves on the measured signal-to-noise ratio, the anechoic termination was set at the end of the duct. The transmission loss of the resonator can be obtained from the three values of upstream and downstream reflection coefficients and the transfer function between two microphones, one upstream and one downstream.[28] The frequency response function between the upstream part and downstream part was measured using flush-mounted $\frac{1}{4}$-in. microphones 1 and 4 (B&K 4135). Both upstream and downstream reflection coefficients were calculated from the measured pressures at arrays of microphones both upstream and downstream. In obtaining each reflection coefficient, three microphones at each side were employed for reducing errors.[29] The transmission loss can be finally obtained as

$$\mathrm{TL}=20\log_{10}\left|\frac{1+R_d}{1+R_u}\right|-20\log_{10}|H_{14}|, \tag{31}$$

where R_d and R_u are the downstream and upstream reflection coefficients and H_{14} means the frequency response function between the microphones 1 and 4.

When a fluid flow is supplied to the duct, the flow-generated turbulent noise is continuously added to the source signal in the whole duct system. The correlation technique was adopted to suppress flow-generated noise from the turbulent flow.[29,30] The sound pressure, measured at a microphone located upstream of the actual measurement sensors, was used as the reference signal for use in the correlation technique. The reference microphone was positioned from the following relation:[30]

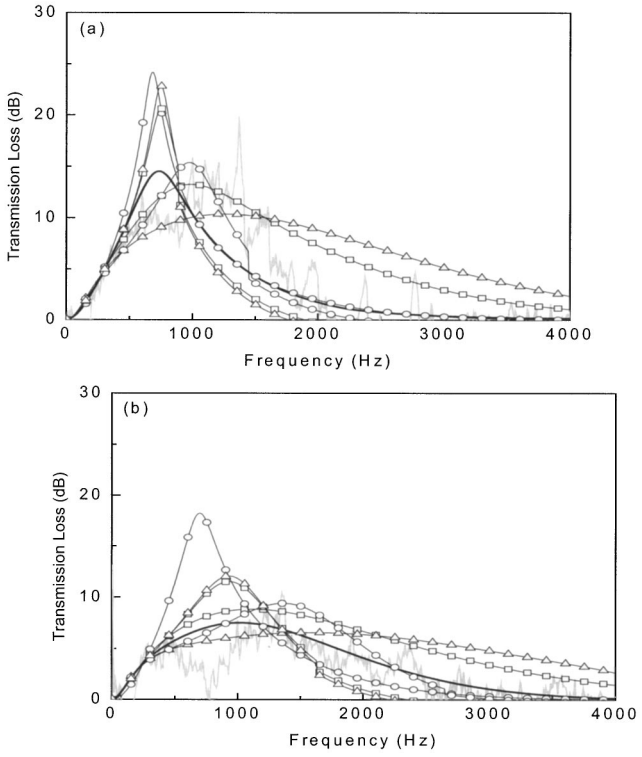


FIG. 19. A comparison of measured and predicted transmission loss of a concentric tube resonator ($l/d_2=0.36$, $\sigma=5\%$): ———, measured; ———, predicted by present model; —■—, predicted by Bauer's model; —●—, predicted by Sullivan's model; —▲—, predicted by Rao and Munjal's model; —□—, predicted by Kooi and Sarin's model; —○—, predicted by Cummings' model; —△—, predicted by Kirby and Cummings' model. (a) M=0.1, (b) M=0.2.

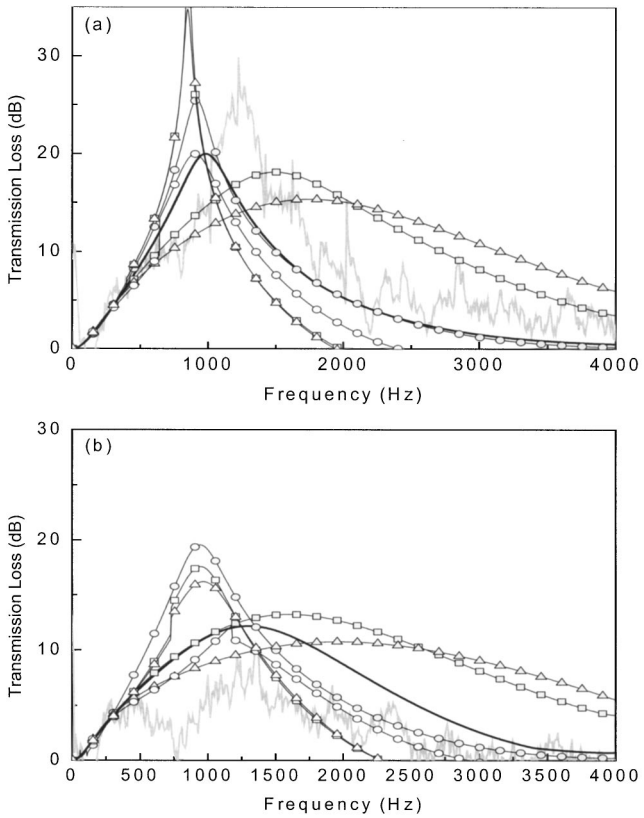


FIG. 20. A comparison of measured and predicted transmission loss of a concentric tube resonator ($l/d_2=0.36$, $\sigma=10\%$): ———, measured; ———, predicted by present model; —■—, predicted by Bauer's model; —●—, predicted by Sullivan's model; —▲—, predicted by Rao and Munjal's model; —□—, predicted by Kooi and Sarin's model; —○—, predicted by Cummings' model; —△—, predicted by Kirby and Cummings' model. (a) M=0.1, (b) M=0.2.

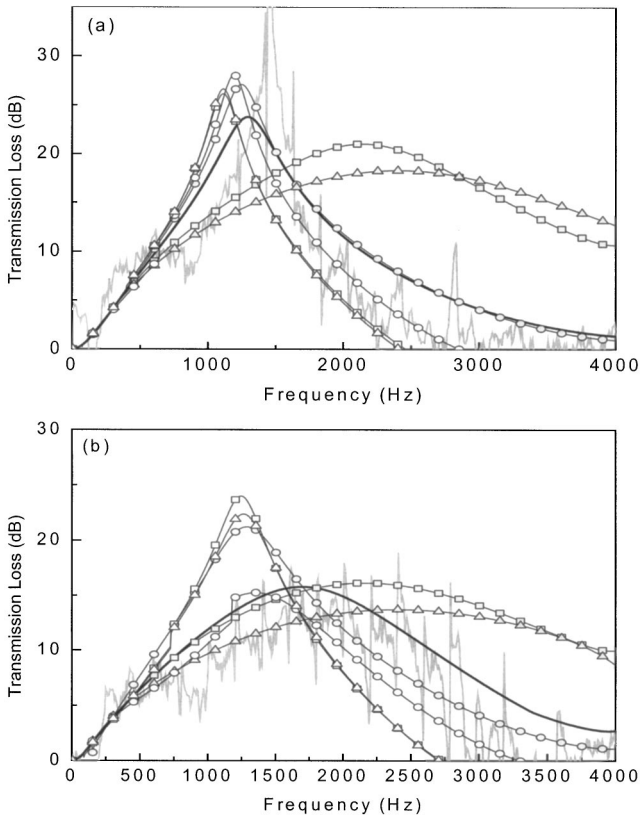


FIG. 21. A comparison of measured and predicted transmission loss of a concentric tube resonator ($l/d_2=0.36$, $\sigma=15\%$): ———, measured; ———, predicted by present model; —■—, predicted by Bauer's model; —●—, predicted by Sullivan's model; —▲—, predicted by Rao and Munjal's model; —□—, predicted by Kooi and Sarin's model; —○—, predicted by Cummings' model; —△—, predicted by Kirby and Cummings' model. (a) M=0.1, (b) M=0.2.

$$ks > 2\pi M. \tag{32}$$

Here, s denotes the distance between the reference microphone and the nearest other microphone, i.e., microphone 3. The number of averaging in obtaining the frequency response function was 500 and the reflection coefficient of the anechoic termination was far less than 0.1 within the valid frequency range.

B. Comparison of measured and predicted transmission loss

The measured transmission losses are compared with the predicted ones, which use the proposed and existing models for the impedance of perforates. The valid frequency range is limited by the cutoff frequency of the first transverse acoustic mode, which corresponds to $kd_2/2=3.83$ [3.8 kHz; (0,1) mode] when the inlet and outlet of the cylindrical expansion chamber are located perfectly on the center of the circular section of the chamber.[31]

Figures 19–21 compare the measured and predicted transmission losses of the short resonator, length to diameter ratio 0.36, with a grazing flow Mach number of 0.1 and 0.2, and a perforate porosity of 5%, 10%, and 15%, respectively. The two figures show a general trend of acoustically short resonators, namely that an increase of porosity raises the primary peak level in the transmission loss curve, while the primary peak becomes smoothed and the effective frequency range is expanded with an increase of Mach number. One can observe that the proposed impedance model yields the best results among all of the predicted curves. When the Bauer's model[9] is employed in the prediction, a slightly different result from that of the present model can be seen and Rao and Munjal's model[17] results in further errors in the prediction. Figures 22–24 compare the measured and predicted transmission losses for the long resonator, length to diameter ratio 1.8, for grazing flow Mach numbers of 0.1 and 0.2, with perforate porosity of 4.7%, 9%, and 14.7%, respectively. The characteristic of the transmission loss curve differs mostly from that of a simple expansion chamber in so far as the third major lobe and the higher frequency performance are concerned. One can see that the proposed impedance model yields the best results among the predicted curves. It is observed that the Sullivan's model[13] produces prediction results nearly as good as the present model, except for the case when both the flow velocity is high and the porosity is low. Predicted transmission losses using Bauer's model[9] and Rao and Munjal's model[17] reveal that these two models can be acceptable only when the flow Mach number is near 0.2 and the prediction is limited to the lobe that rep-

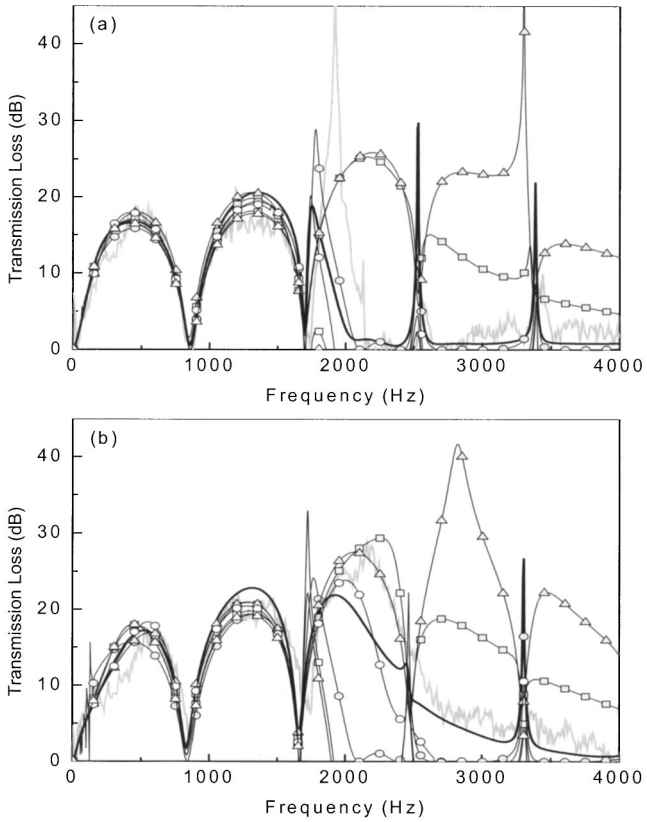


FIG. 22. A comparison of measured and predicted transmission loss of a concentric tube resonator ($l/d_2=1.8$, $\sigma=4.7\%$): ———, measured; ———, predicted by present model; —■—, predicted by Bauer's model; —●—, predicted by Sullivan's model; —▲—, predicted by Rao and Munjal's model; —□—, predicted by Kooi and Sarin's model; —○—, predicted by Cummings' model; —△—, predicted by Kirby and Cummings' model. (a) M=0.1, (b) M=0.2.

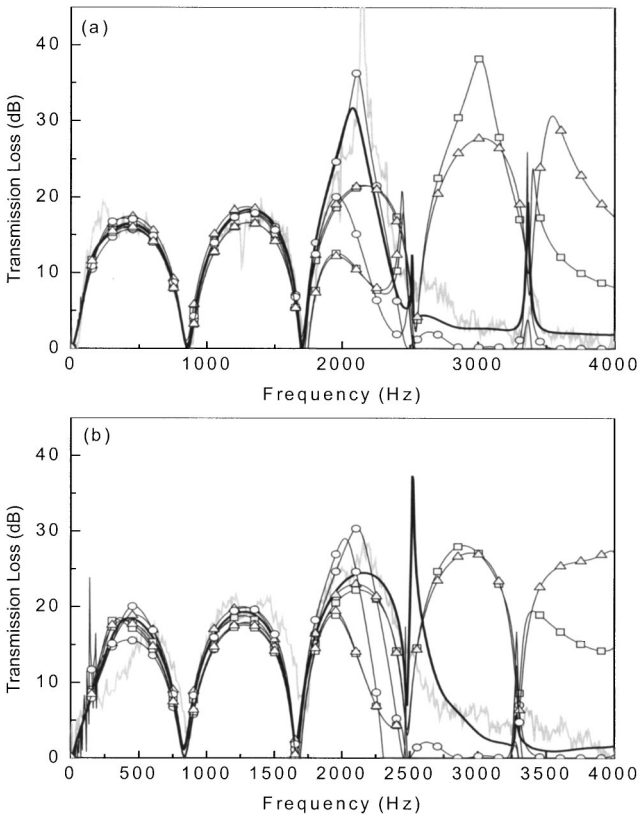


FIG. 23. A comparison of measured and predicted transmission loss of a concentric tube resonator ($l/d_2=1.8$, $\sigma=9\%$): ———, measured; ———, predicted by present model; —■—, predicted by Bauer's model; —●—, predicted by Sullivan's model; —▲—, predicted by Rao and Munjal's model; —□—, predicted by Kooi and Sarin's model; —○—, predicted by Cummings' model; —△—, predicted by Kirby and Cummings' model. (a) M=0.1, (b) M=0.2.

resents the first resonator peak, i.e., the third major lobe in Figs. 22–24.

It should be noted that very-narrow-band fluctuation in the measured transmission loss is not of concern when the predicted data is compared with experimental one. Furthermore, if the fluctuation in the data is owing to the aerodynamic fluctuation, it is not related with the actual acoustic event acting on the holes. The actual sound activity is related with the trend of the curve. What is really talked about is the difference in spectral trends. Notwithstanding the fact that the difference between predictions is significant at high frequencies for the acoustically long chamber ($L/D > 0.41-0.43$), one can still observe small differences in predictions at low frequencies. When the comparison plots for the acoustically short chambers ($L/D < 0.41-0.43$), the improvement of prediction accuracy can be very clearly observed and it is significant. Even in the low-frequency range, one can observe some difference between predicted results and the smallest difference at about 500 Hz is 1–2 dB in the examples of this paper. Every muffler designer knows that the reduction of exhaust noise by 1–2 dB at low frequencies is quite a difficult task in refining an existing muffler. One may recall that the acoustically short chambers are used very often in actual mufflers of vehicles due to their limited geometrical space. A muffler is designed as a combination of several small modules. By combining several modules, one can obtain high sound attenuation at mid to high frequencies in general, but usually a dip in TL appears in the very-low-frequency region, which is very hard to compensate. In this case, 1–2 dB difference in prediction is very important in the design of mufflers. In addition, it is true that the low-frequency performance is important in evaluating the overall acoustical performance of silencers, but that is not enough in recent days. Because the noises from various sources are reduced very much in modern cars, the importance of sound quality, in particular for interior noise, is ever increasing and, accordingly, the mid- to high-frequency sounds are considered more and more important. Needless to say, this is also true for noise radiated from intake and exhaust systems.

IV. CONCLUSIONS

It was intended to propose a practical and precise empirical impedance model for a perforate under grazing flow conditions. For this purpose, the acoustic impedance of perforated elements under grazing mean flow conditions was measured when considering the effects of the following parameters: orifice diameter, orifice thickness, perforation ratio, mean flow velocity, and frequency. The range of involved parameters cover the practical data span of perforates in typi-

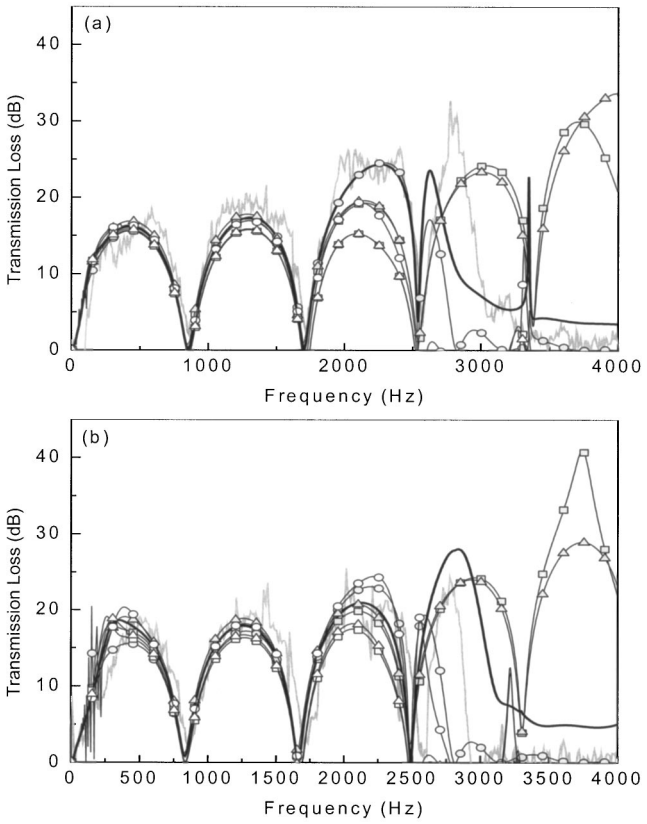


FIG. 24. A comparison of measured and predicted transmission loss of a concentric tube resonator ($l/d_2=1.8$, $\sigma=14.7\%$): ———, measured; ———, predicted by present model; —■—, predicted by Bauer's model; —●—, predicted by Sullivan's model; —▲—, predicted by Rao and Munjal's model; —□—, predicted by Kooi and Sarin's model; —○—, predicted by Cummings' model; —△—, predicted by Kirby and Cummings' model. (a) M=0.1, (b) M=0.2.

cal exhaust systems of internal combustion engines such that $60 \leqslant f \leqslant 4000$ Hz, $0 \leqslant M \leqslant 0.2$, $2 \leqslant d \leqslant 9$ mm, $1 \leqslant t \leqslant 5$ mm, and $2.79 \leqslant \sigma \leqslant 22.3\%$. It is noted that these applicable ranges of perforate parameters are far extended compared to the previous models. The nonlinear regression technique was used for deriving the empirical impedance model of which high correlation coefficients larger than about 0.85 were obtained in general, which assures the validity of the proposed impedance model.

It is found that the magnitude of the resistance increases in general as the mean flow Mach number increases, but the rate of decrease of resistance with frequency is nearly the same for every Mach number condition. However, the decreasing trend of resistance with increase of frequency changes to an increasing one above a certain frequency, at which the resistance is nearly zero. Reactance increases with frequency and its magnitude generally decreases as the flow Mach number increases, although the measured reactance values are nearly the same at high frequencies, when the Mach number is smaller than about 0.1. As far as the porosity is concerned, the resistance shows V-shaped curves for all porosities centered at around 2.4 kHz, where the resistance is very small, near to zero, regardless of the porosity. With increasing porosity, both resistance and reactance decrease in general. Both resistance and reactance increase in general with an increase of orifice diameter, but at low frequencies the resistance decreases with increasing the frequency and then increases after passing a certain frequency. In general, resistance decreases and reactance increases with an increase of the orifice thickness. The overall trend of the resistance is like a V-shape and the resistance minimum is at about 2.4 kHz.

The empirical model was obtained by using a nonlinear regression analysis on the various results of the parametric tests. The proposed empirical model for orifice impedance, with a very high correlation coefficient, is applied to the prediction of the transmission loss of four concentric resonators, which are typical geometric configurations for both acoustically short and long through-flow resonators. It is observed that the proposed impedance model yields the best results among other predicted curves employing earlier impedance models.

ACKNOWLEDGMENTS

This work was partially supported by the BK21 Project and NRL. The authors would like to thank Dr. Keith S. Peat at Loughborough University, U.K., for helpful discussions about this work.

[1] J. W. S. Rayleigh, *Theory of Sound* (Dover, New York, 1945), revised ed., Vol. II, Chap. XIX.
[2] I. B. Crandall, *Theory of Vibrating Systems and Sound* (Van Nostrand, New York, 1927), pp. 229–241.
[3] L. J. Sivian, "Acoustic impedance of small orifices," J. Acoust. Soc. Am. **7**, 94–101 (1935).
[4] U. Ingard, "On the theory and design of acoustic resonator," J. Acoust. Soc. Am. **25**, 1037–1062 (1953).
[5] T. H. Melling, "The acoustic impedance of perforates at medium and high sound pressure levels," J. Sound Vib. **29**, 1–65 (1973).
[6] V. A. Fok, *Doklady Akademii nauk SSSR 31* (1941, in Russian) Alternatively, see S. N. Rschevkin, *A Course of Lectures on the Theory of Sound* (Pergamon, London, 1963).
[7] J. W. Sullivan and M. J. Crocker, "Analysis of concentric-tube resonators having unpartitioned cavities," J. Acoust. Soc. Am. **64**, 207–215 (1978).
[8] G. D. Garrison, "Suppression of combustion oscillations with mechanical damping devices," Pratt and Whitney Aircraft Rep. PWA FR3299 (1969).
[9] D. Ronneberger, "The acoustic impedance of holes in the wall of flow ducts," J. Sound Vib. **24**, 133–150 (1972).
[10] P. D. Dean, "An in situ method of wall acoustic impedance measurements in flow duct," J. Sound Vib. **34**, 97–130 (1974).
[11] A. B. Bauer, "Impedance theory and measurements on porous acoustic liners," J. Aircr. **14**, 720–728 (1977).
[12] A. S. Hersh, B. Walker, and M. Bucka, "Effect of the grazing flow on the acoustic impedance of Helmholtz resonators consisting of single and clustered orifices," AIAA Pap. , 78–1124 (1978).
[13] J. W. Sullivan, "A method for modeling perforated tube muffler components. II. Application," J. Acoust. Soc. Am. **66**, 779–788 (1979).
[14] K. Jayaraman and K. Yam, "Decoupling approach to modeling perforated tube muffler components," J. Acoust. Soc. Am. **69**, 390–396 (1981).
[15] A. Goldman and R. L. Panton, "Measurement of the acoustic impedance of an orifice under a turbulent boundary layer," J. Acoust. Soc. Am. **60**, 1397–1404 (1976).
[16] A. Goldman and C. H. Chung, "Impedance of an orifice under a turbulent boundary layer with pressure gradient," J. Acoust. Soc. Am. **71**, 573–579 (1982).
[17] K. N. Rao and M. L. Munjal, "Experimental evaluation of impedance of perforate with grazing flow," J. Sound Vib. **108**, 283–295 (1986).
[18] J. W. Kooi and S. L. Sarin, "An experimental study of the acoustic impedance of Helmholtz resonator arrays under a turbulent boundary layer," AIAA Pap. , 81–1998 (1981).
[19] A. Cummings, "The effects of grazing turbulent pipe-flow on the impedance of an orifice," Acustica **61**, 233–242 (1986).

[20] R. Kirby and A. Cummings, ''The impedance of perforated plates subjected to grazing gas flow and backed by porous media,'' J. Sound Vib. **217**, 619–636 (1998).
[21] U. Ingard and H. Ising, ''Acoustic nonlinearity of an orifice,'' J. Acoust. Soc. Am. **42**, 6–17 (1967).
[22] H. Fujita, ''Turbulent flow in smooth and rough walled square ducts,'' Trans. Jpn. Soc. Mech. Eng., Ser. B **45**, 197–207 (1979).
[23] J. W. Sullivan, ''A method for modeling perforated tube muffler components. I. Theory,'' J. Acoust. Soc. Am. **66**, 772–778 (1979).
[24] M. L. Munjal, K. N. Rao, and A. D. Sahasrabudhe, ''Aeroacoustic analysis of perforated muffler components,'' J. Sound Vib. **114**, 173–188 (1987).
[25] K. S. Peat, ''A numerical decoupling analysis of perforated pipe silencer elements,'' J. Sound Vib. **123**, 199–212 (1988).
[26] P. T. Thawani and K. Jayaraman, ''Modeling and applications of straight-through resonators,'' J. Acoust. Soc. Am. **73**, 1387–1389 (1983).
[27] J.-G. Ih and B.-H. Lee, ''Implication of geometric factors of simple expansion chambers for their acoustic performance,'' in Proceedings of the International Noise and Vibration Control Conference (Noise '93) May 1993, St. Petersburg, Russia, Vol. 3, pp. 155–160.
[28] J. Y. Chung and D. A. Blaser, ''Transfer function method of measuring in-duct acoustic properties. I. Theory,'' J. Acoust. Soc. Am. **68**, 907–913 (1980).
[29] S.-H. Jang and J.-G. Ih, ''On the multiple microphone method for measuring in-duct acoustic properties in the presence of mean flow,'' J.Acoust. Soc. Am. **103**, 1520–1526 (1998).
[30] M. Abom and H. Boden, ''Error analysis of two-microphone measurements in ducts with flow,'' J. Acoust. Soc. Am. **83**, 2429–2438 (1988).
[31] J.-G. Ih and B.-H. Lee, ''Analysis of higher order mode effects in the circular expansion chamber with mean flow,'' J. Acoust. Soc. Am. **77**, 1377–1388 (1985).

The shock formation distance in a bounded sound beam of finite amplitude

Chao Tao, Jian Ma, Zhemin Zhu, and Gonghuan Du[a)]
State Key Lab of Modern Acoustics & Institute of Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

Zihong Ping
Hangzhou Applied Acoustics Institute, Hangzhou Fuyang 311400, People's Republic of China

(Received 10 June 2002; revised 4 April 2003; accepted 14 April 2003)

This paper investigates the shock formation distance in a bounded sound beam of finite amplitude by solving the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation using frequency-domain numerical method. Simulation results reveal that, besides the nonlinearity and absorption, the diffraction is another important factor that affects the shock formation of a bounded sound beam. More detailed discussions of the shock formation in a bounded sound beam, such as the waveform of sound pressure and the spatial distribution of shock formation, are also presented and compared for different parameters. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1579002]

PACS numbers: 43.25.Cb, 43.25.Jh [MFH]

I. INTRODUCTION

Nonlinear characteristics of a one-dimensional acoustical field, such as plane, spherical, and cylindrical wave, have been solved analytically.[1] However, most of the bounded sound beam, such as the piston field, which is widely used in practical application, cannot be solved analytically, and numerical simulation is necessary. In the last two decades, several computational algorithms have been developed.[2–12] Additionally, the nonlinear characteristics of an acoustical field are discussed widely by using these algorithms, a series of problems has been solved successfully.[2–10,13] An overview of the modern computational techniques can be found in a recent monograph on nonlinear acoustics.[14] Because of the restriction of huge computational time, one typical and important nonlinear characteristic—shock formation—was seldom researched in detail in the early work until some fast algorithms were developed.[13,15,16] Recently, it has been discussed widely.[17–23] In particular, Khokhlova *et al.* presented a fast algorithm to reduce computational time and discuss the shock formation in the near field of a cw plane piston source by using this algorithm.[23]

As is known, the shock formation has many particular characteristics.[1,12] It is important to evaluate the shock formation distance and plot out the shock formation area in a finite amplitude acoustical field for many theoretical and practical problems. Unfortunately, the current nonlinear theory only gives the shock formation distance of one-dimensional wave in lossless media.[1] Therefore, in order to give insight into the shock formation of a sound beam, in this paper, we discuss the shock formation distance in the finite amplitude field radiated from an intense piston source by solving the KZK equation using Fourier series expansion and numerical method. Several factors, such as diffraction, absorption, and source intensity, are considered to investigate the intense nonlinear acoustic field. Study will indicate that the shock formation in the piston field is much more complex than that in the one-dimensional field. Besides the nonlinearity and absorption, the diffraction is another major role that influences the shock formation in a bounded sound beam. To describe the influence of diffraction, the relation between the Rayleigh distance and the shock formation distance at the axis of a piston field is given. Finally, the waveform of pressure and the spatial distribution of shock formation are also discussed.

This paper is organized as follows: In the second section, the algorithm is described briefly and the numerical method of evaluating the shock formation distance is proposed. In the third section, calculation results are presented and discussed. In the last section, we draw the conclusion.

II. NUMERICAL MODEL

A. Model equations

Our study is based on the numerical solution of the KZK equation:[24,25]

$$\left(4\frac{\partial^2}{\partial\tau\partial\sigma}-\overline{\nabla}_{\perp}^2-4\alpha r_0\frac{\partial^3}{\partial\tau^3}\right)p=2\frac{1}{\sigma_D}\frac{\partial^2}{\partial\tau^2}p^2. \tag{1}$$

Here $\tau=\omega(t-z/c_0)$, $\sigma=z/r_0$, $\xi=\rho/a$, $p=(P-P_0)/\rho_0 c_0 u_0$, P_0 and ρ_0 are ambient pressure and ambient density, P is the pressure, t is the time, z and ρ are the coordinate along the axial and radial direction respectively. $\overline{\nabla}_{\perp}^2=(\partial^2/\partial\xi^2)+(1/\xi)(\partial/\partial\xi)$ is the nondimensional transverse Laplace operator with respect to ξ. Further, ω^{-1} is a characteristic time and a is a characteristic length transverse to the direction of propagation. For an axisymmetric piston sinusoidal source, $\omega/2\pi$ is frequency and a is source radius, u_0 and c_0 are the characteristic value of the velocity and isentropic sound speed, respectively. $M=u_0/c_0$ is the Mach number; r_0 is the Rayleigh distance; $\alpha=D\omega^2/2c_0^3$ is the absorption coefficient, D being the diffusivity of sound; and $\sigma_D=l_D/r_0$, where $l_D=c_0/\beta M\omega$ is the shock formation dis-

a)Electronic mail: zhudu@nju.edu.cn

 0001-4966/2003/114(1)/114/8/$19.00

tance of a plane wave in lossless media. The boundary condition for Eq. (1) is $p(\sigma=0,\xi,\tau)=f(\xi,\tau)$, where $u_0 f(\xi,\tau)$ is the normal velocity distribution in the plane of source.

When the source is periodic in time with period $2\pi/\omega$, the solution of Eq. (1) can be expanded by using the Fourier series:

$$p=\sum_{n=1}^{\infty}(g_n \sin n\tau + h_n \cos n\tau), \tag{2}$$

and its coefficient g_n and h_n can be obtained by solving the following equations:

$$\begin{aligned}\frac{\partial g_n}{\partial \sigma} &= -n^2 A g_n + \frac{1}{4n}\overline{\nabla}_{\perp}^2 h_n + N\frac{n}{2}\left(\frac{1}{2}\sum_{p=1}^{n-1}(g_p g_{n-p} - h_p h_{n-p}) - \sum_{p=n+1}^{\infty}(g_{p-n}g_p + h_{p-n}h_p)\right),\\ \frac{\partial h_n}{\partial \sigma} &= -n^2 A h_n - \frac{1}{4n}\overline{\nabla}_{\perp}^2 g_n + N\frac{n}{2}\left(\frac{1}{2}\sum_{p=1}^{n-1}(h_p g_{n-p} + g_p h_{n-p}) + \sum_{p=n+1}^{\infty}(h_{p-n}g_p - g_{p-n}h_p)\right).\end{aligned} \tag{3}$$

Here n is the harmonic number. The nonlinearity parameter N and the absorption parameter A are defined as

$$A=\alpha r_0 \quad \text{and} \quad N=r_0/l_D=1/\sigma_D.$$

Further, we define the nonlinearity absorption ratio as

$$N/A=1/\alpha l_D.$$

The boundary condition for g_n and h_n at $\sigma=0$ is

$$g_n=\frac{1}{\pi}\int_{-\pi}^{\pi} f \sin n\tau \, d\tau \quad \text{and} \quad h_n=\frac{1}{\pi}\int_{-\pi}^{\pi} f \cos n\tau \, d\tau. \tag{4}$$

For the case of an axisymmetric piston source which oscillates sinusoidally, we have the boundary condition

$$\left.\begin{aligned} &g_1=1\,|\xi|\leqslant 1, \quad g_1=0\,|\xi|>1\\ &g_2=g_3=\cdots=0\\ &h_1=h_2=\cdots=0\end{aligned}\right\}, \quad \text{when} \quad \sigma=0.$$

Equations (3) have been solved numerically.[2,3] To reduce the calculation time, we adopt and improve the calculation technique presented in Ref. 23. The parameter values of operation are chosen as follows: The radial and axial integral steps are equal to $\Delta\xi=5\times10^{-3}$ and $\Delta\sigma=2.5\times10^{-5}$, respectively. The initial number of harmonics is $n=40$, and the number of harmonics is increased by $\Delta n=100$ when the absolute of the last harmonic amplitude exceeds the threshold 2×10^{-4}. The size of the spatial window in the radial direction $\xi_{\max}^{(n)}$ is chosen depending on the harmonics number n and the propagation distance σ,

$$\xi_{\max}^{(n)}=\begin{cases}\xi_{\max}\times 1, & 1\leqslant n\leqslant 30,\\ \xi_{\max}\times 0.8, & 31\leqslant n\leqslant 100,\\ \xi_{\max}\times 0.6, & 101\leqslant n\leqslant 300,\\ \xi_{\max}\times 0.4, & 301\leqslant n\leqslant 800,\\ \xi_{\max}\times 0.22, & 801\leqslant n\leqslant 1000,\end{cases}$$

$$\text{here} \quad \xi_{\max}=\begin{cases}5{,}0\leqslant\sigma\leqslant 0.3,\\ 7.5{,}0.3\leqslant\sigma\leqslant 0.5,\\ 10{,}0.5\leqslant\sigma\leqslant 0.8.\end{cases}$$

To improve computational speed more, we write the entire source code in C++ language, compile and optimize aiming at Pentium IV 1.6 G CPU by using Intel C++ Compiler 5.0.1.

B. Numerical evaluation of shock formation distance

The analytical nonlinear theory evaluates the shock formation distance of a plane wave in lossless media according to a multivalueness beginning to appear in the profile of p[1]; in other words when the distance approaches $\sigma=\sigma_D$, the maximum derivative of sound pressure approaches infinity:

$$\max_{0\leqslant\tau\leqslant 2\pi}\left(\frac{dp(\sigma_D,\tau)}{d\tau}\right)=\infty.$$

Since the numerical results cannot express an infinite value, this method cannot be used to evaluate the shock formation distance of the piston field. However, it is enlightening that the maximum derivative of sound pressure can be used to describe the distortion of the waveform. Further, the changes in the wave amplitude due to diffraction, beam converging, etc., may also influence the maximum derivative of sound pressure. We therefore use the maximum derivative of the waveform with normalized peak-to-peak pressure to describe the distortion of the waveform:

$$S_m=\max_{0\leqslant\tau\leqslant 2\pi}\left(\frac{d}{d\tau}\left(\frac{p(\sigma,\tau)}{p_{pp}}\right)\right), \tag{5}$$

where p_{pp} is the peak-to-peak value of pressure p. Because the waveform p is normalized in Eq. (5), this equation can correctly describe the distortion of the waveform even when its amplitude is changing and the changes in waveform amplitude will not influence the judgment of the shock formation distance. The larger S_m indicates that the waveform is distorted more seriously. Then, we use a threshold ϵ to judge whether the shock has formed; i.e., $S_m\geqslant\epsilon$ indicates that the shock has formed, otherwise the shock has not formed. The next problem is how to choose the threshold ϵ. We notice that the piston field approaches a plane field when r_0 approaches infinity. Therefore, the shock formation distance of the piston field evaluated by this threshold value should also approach the plane wave's distance when $r_0\to\infty$. According to the above discussion, we evaluate the shock formation distance of the piston field as follows:

First, an auxiliary plane wave model is used to determine the threshold ϵ. For the given sinusoidal wave amplitude at the source, the plane wave propagating in lossy media

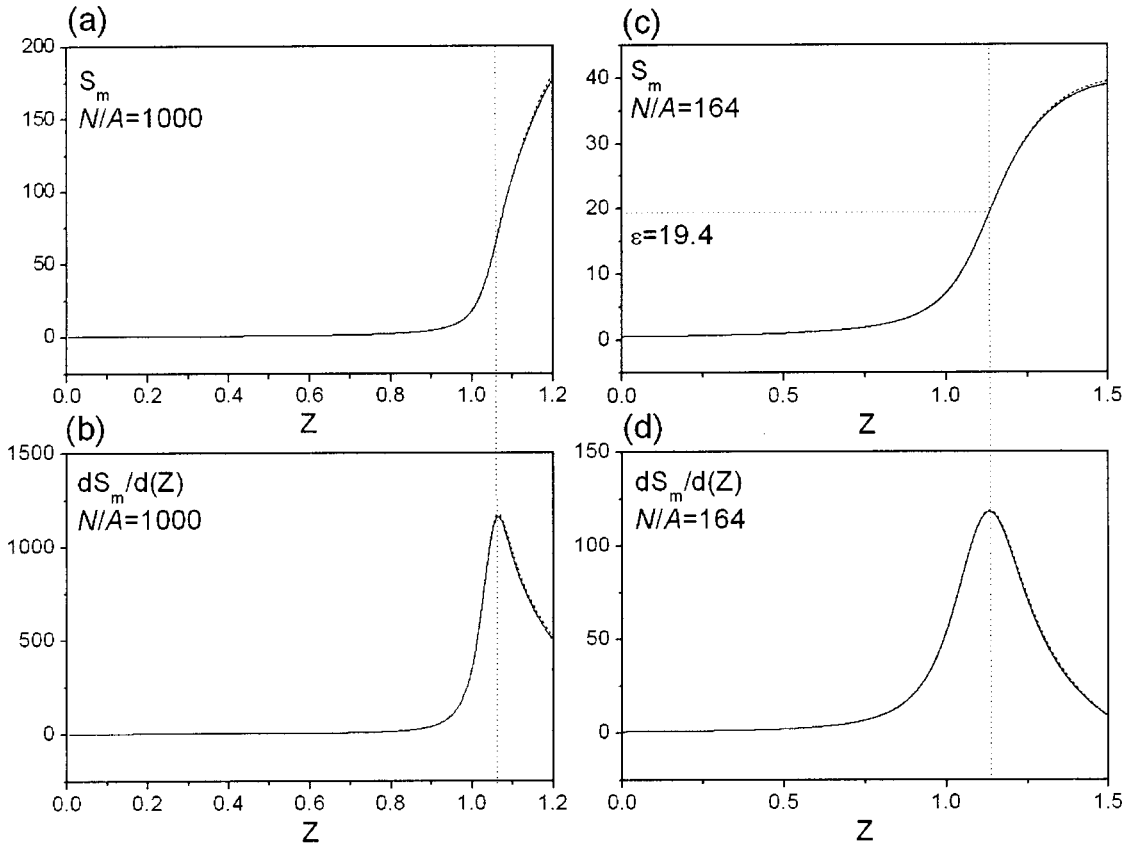


FIG. 1. Determine the threshold ϵ. In (a) and (b), the solid line and dashed line are calculated with 3200 harmonics and 5000 harmonics. In (c) and (d), the solid line and dashed line are calculated with 500 harmonics and 1000 harmonics. (a) The curve of S_m vs Z with $N/A=1000$. (b) The derivative of S_m with $N/A=1000$. (c) The curve of S_m vs Z with $N/A=164$. (d) The derivative of S_m with $N/A=164$.

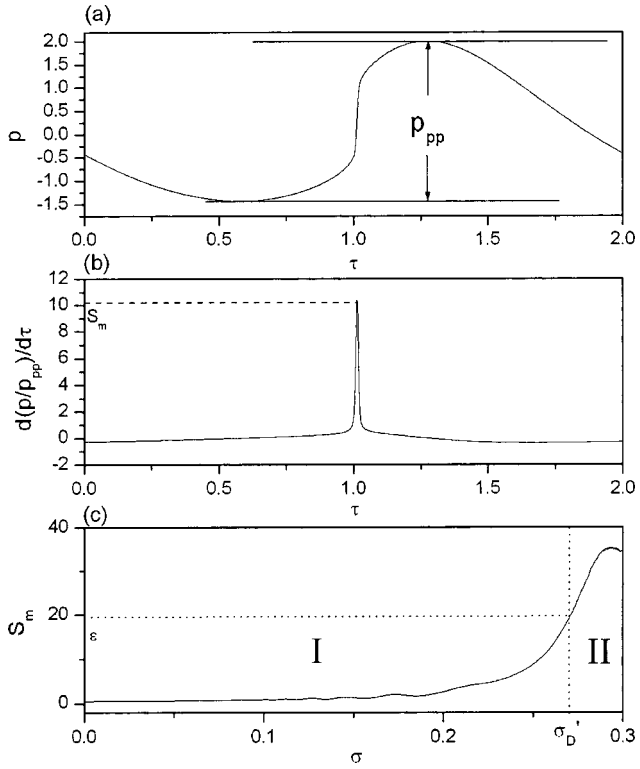


FIG. 2. Steps of evaluating the shock formation distance of a piston field. (a) Compose the waveform of sound pressure with $p=\Sigma_{n=1}^{\infty}(g_n \sin n\tau + h_n \cos n\tau)$. (b) Calculate the derivative $(d/d\tau)[p(\tau)/p_{pp}]$. (c) Draw the curve of S_m vs σ, here in region II, wave is in the shock formation.

is solved by using the following equation with the same value of $N/A(N/A \gg 1)$ as that in the KZK equation:

$$\left(2\frac{\partial}{\partial Z}-2\frac{A}{N}\frac{\partial^2}{\partial \tau^2}\right)p=\frac{\partial}{\partial \tau}p^2, \tag{6}$$

where $Z=z/l_D$. Clearly, this equation is obtained by ignoring the diffractive term of Eq. (1) and it is an equivalent of Burgers equation. Therefore, its solution can be obtained by solving the coupling partial differential Eqs. (3) with ignored diffractive term.

Then S_m [Figs. 1(a) and (c)] and its spatial-derivative dS_m/dZ [Figs. 1(b) and (d)] over the propagation distance are calculated as a function of Z. The distance $Z^{(N/A)}$ is found where the function dS_m/dZ reaches a maximum. At this distance $Z^{(N/A)}(z=l_D^{(N/A)})$, the corresponding value of $\epsilon=S_m(Z^{(N/A)})$ is evaluated and used as a threshold for evaluating the shock formation in the full diffraction problem. Also, the distance $l_D^{(N/A)}=Z^{(N/A)}\times l_D$ is considered the shock formation distance of a plane wave propagating in a lossy media $(1 \ll N/A < \infty)$. For instance, in the lossy media with $\alpha=0.025\,33\ \mathrm{m}^{-1}$, $l_D=0.24$ m and $N/A=164$, we have

TABLE I. The value of $Z^{(N/A)}$ with different N/A.

N/A	100	164	395	1000	2000	$\rightarrow\infty$
$Z^{(N/A)}$	1.158	1.1333	1.091	1.060	1.039	$\rightarrow$1.000

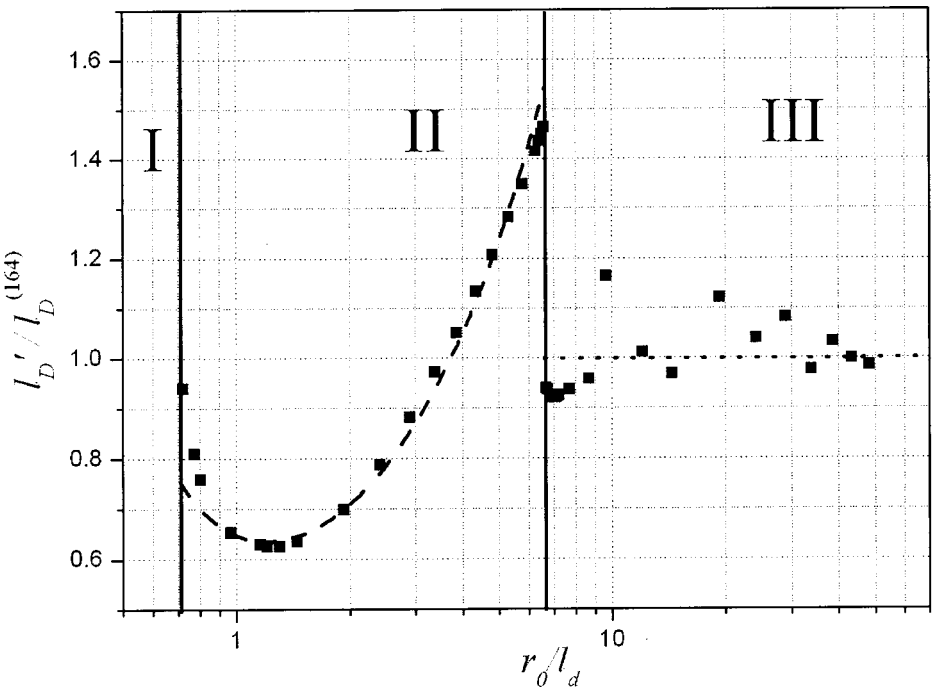


FIG. 3. The shock formation distance at the axis of a piston field.

ϵ=19.4, $Z^{(164)}$=1.133, and $l_D^{(164)}$=0.272 m. Figure 1 illustrates the steps determining the threshold ϵ. It also shows that the curves calculated with different harmonic numbers are almost identical when the harmonic number is larger than $\pi N/A$.[1,23]

Finally, the bounded sound beam is solved by using the KZK equation. When the S_m predicted by the KZK equation exceeds the value of threshold ϵ, it is considered that the shock has formed. The shock formation distance is the nearest position where $S_m > \epsilon$ (Fig. 2).

One benefit of the above algorithm is that it is compatible with the analytical theory for a strong nonlinear wave ($N/A \gg 1$). We find the increase of *N/A* causes the decrease of $Z^{(N/A)}$ (Fig. 1 and Table I). When the value of *N/A* approaches ∞, the value of $Z^{(N/A)}$ will be close to 1 ($Z^{(N/A)} \rightarrow l_D$), which implies the shock formation distance evaluated

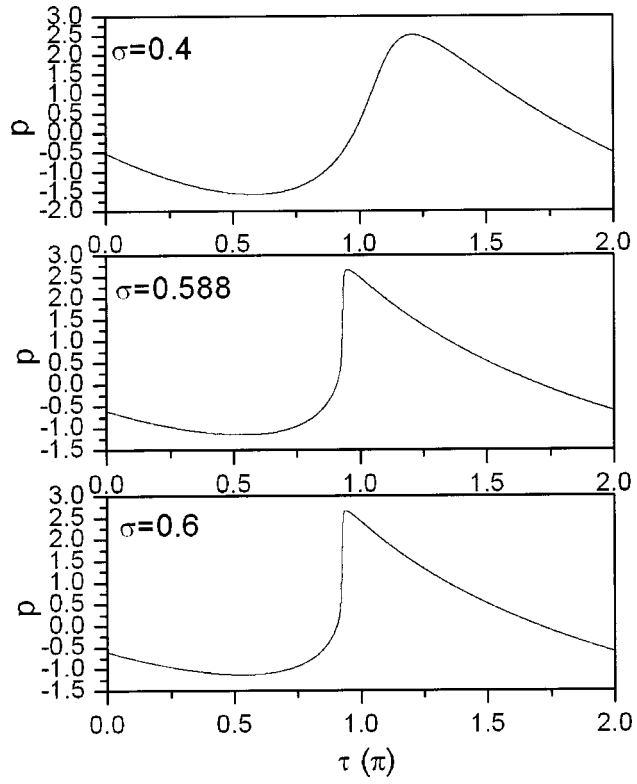


FIG. 4. The waveform of sound pressure in various distances with N = 1.204 and A = 0.007 32.

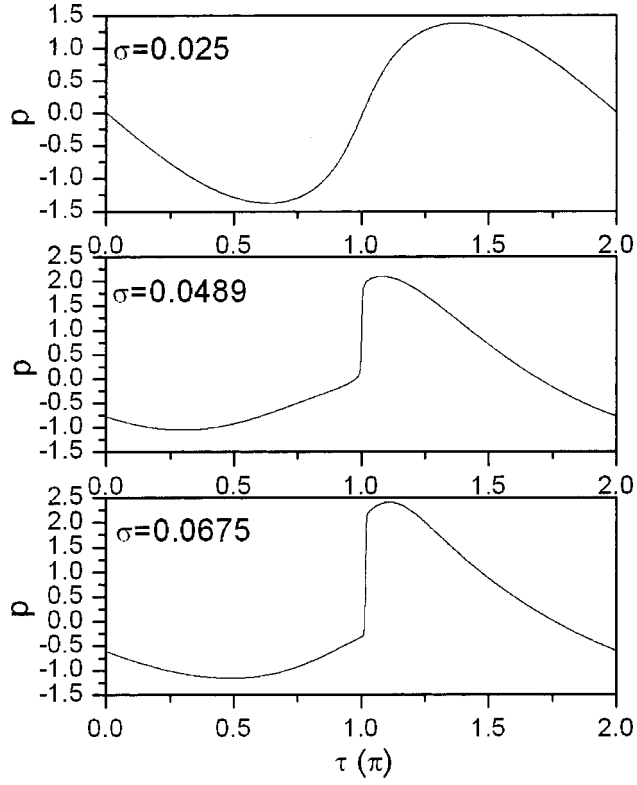


FIG. 5. The waveform of sound pressure in various distances with N = 24.083 and A = 0.146 41.

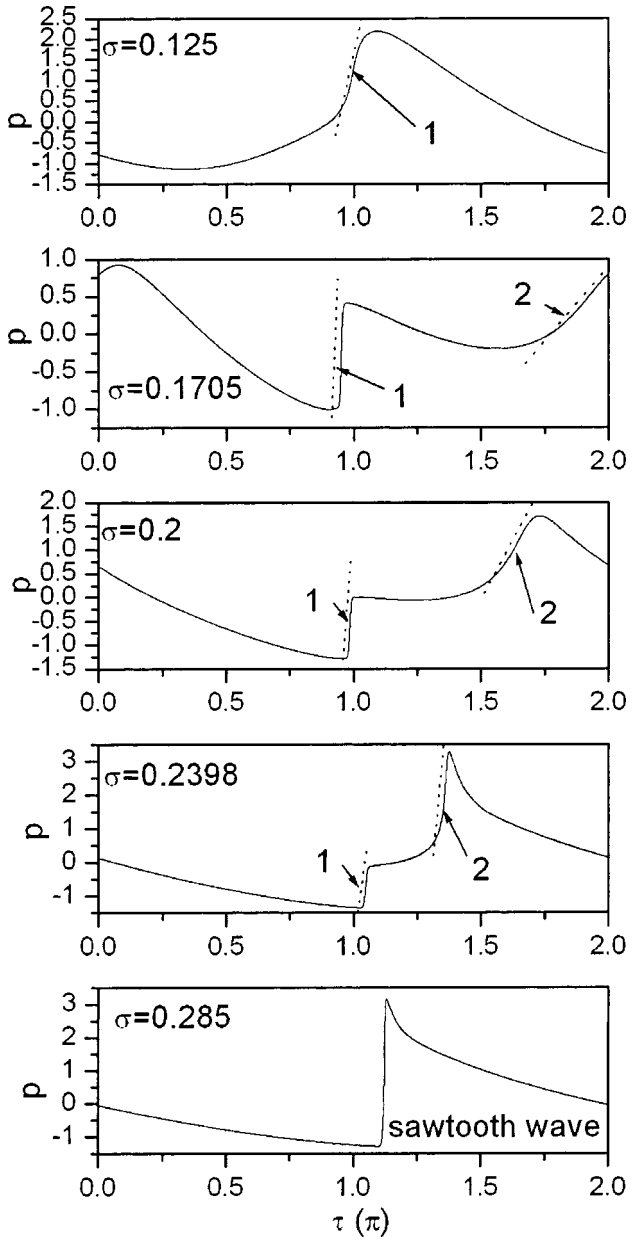


FIG. 6. The growing process of the plane-wave-like shock and spherical-wave-like shock, where shock 1 is the plane-wave-like shock and shock 2 is the spherical-wave-like shock. The parameters are $A=0.0425$, $N=6.984$.

with the above algorithm approaches the analytical solution l_D in the case of zero absorption $\alpha\to 0$. And when N/A is large enough ($N/A\gg 1$), the difference between $l_D^{(N/A)}$ and l_D is small. These meet the requirement of the above discussion.

Another important reason is that it is of benefit to evaluate the shock formation distance exactly enough even when ϵ has a slight error, because ϵ is set as the value $S_m(Z^{(N/A)})$ where the curve of the maximum slope is steepest. For example, in the lossy media with $N/A=164$, a 5% error of ϵ only brings a 0.7% error to the shock formation distance.

III. RESULT AND DISCUSSION

Based on the above model equations and numerical evaluation of shock formation distance, we study the shock formation distance of a piston field in this section. The piston field with a total of 40 sets of parameters N and A is simulated. All the fields have the same $N/A=164$. Therefore, we estimate the shock formation distances l_D' at the axis with the threshold $\epsilon=19.4$. Because $(N/A)^{-1}=\alpha r_0\times(l_D/r_0)$ is identical, we can consider that l_D and α are fixed but r_0 is variable. In other words, it can be considered that the sound beams are all radiated from the source with the frequency 1 MHz and the initial pressure amplitude 0.64 MPa, and propagate in the same media (in water, $\rho_0=1000$ kg/m^3, $\beta=3.5$, $c_0=1500$ m/s, $\alpha=0.025\,33$ m^{-1} at 1 MHz, $l_D=0.24$ m), but the source radius in different $a\in(0.9099$ cm, 7.429 cm) correspond to $r_0\in(0.1734$ m, 11.56 m). We therefore get the relation between $l_D'/l_D^{(164)}$ and r_0/l_D according to our simulation results. The curve of $l_D'/l_D^{(164)}$ vs r_0/l_D is given in Fig. 3. Clearly, the influence of r_0 on l_D' is very complex and Fig. 3 can be divided into three regions. In region III ($r_0/l_D>6.6$), $l_D'/l_D^{(164)}$ is undulate around 1 and converges at 1 for $r_0/l_D\to\infty$, which means $l_D'\to l_D^{(164)}$. This reasonable result can be explained as follows.

When $r_0\gg l_D$, the shock of a piston field forms in the nearfield and, according to the theory of a piston beam, in the near field of a piston field, its spatial-average characteristic is similar to a plane wave. So, its shock formation distance should also approach that of its corresponding plane wave just as our simulation results have shown. The agreement also proves that our numerical simulation and numerical evaluation of shock formation distance is effective to research the shock formation distance of a piston field.

In region II ($0.723<r_0/l_D<6.6$), the shock formation distance increases suddenly at about $r_0/l_D=6.6$, and then decreases with the decrease of r_0/l_D until $r_0/l_D=1.204$, when $0.723<r_0/l_D<1.204$ the shock formation distance increases again. Finally, in region I ($r_0/l_0<0.723$), the shock wave cannot happen at all. To describe the relation between r_0 and l_D', we fit the curve in region II with the following formula since the far-field characteristic of a piston field is similar to a spherical field and the shock forms in the far-field:

$$\frac{l_D'}{l_D^{(164)}}=C_1 r_0\,\exp\left(C_2\frac{l_D}{r_0}\right). \tag{7}$$

Numerical simulation determines the parameters $C_1=0.81$ and $C_2=1.204$; the curve of $l_D'/l_D^{(164)}$ vs r_0/l_D calculated with formula (7) is also plotted in Fig. 3 with the dashed line. It is seen that formula (7) can actually approach our calculation results, which implies that the shock formation distance in region II obeys a similar rule to that of a spherical wave. Although this formula is fitted with $N/A=164$, it is also suitable for the condition that $N/A\to\infty$ after altering the parameters C_1, C_2 slightly.

For an insight into the shock wave in regions II and III, we give the pressure waveform at various distances from source with two sets of typical parameters N and A. Figures 4 and 5 illustrate the pressure waveform with parameters $N=1.204$, $A=0.007\,32$ (region II), and $N=24.083$, $A=0.146\,41$ (region III). Their shocks form at $\sigma=0.588$ and $\sigma=0.0489$, respectively. Clearly, the waveforms are very different. The pressure waveforms in region II are steep and

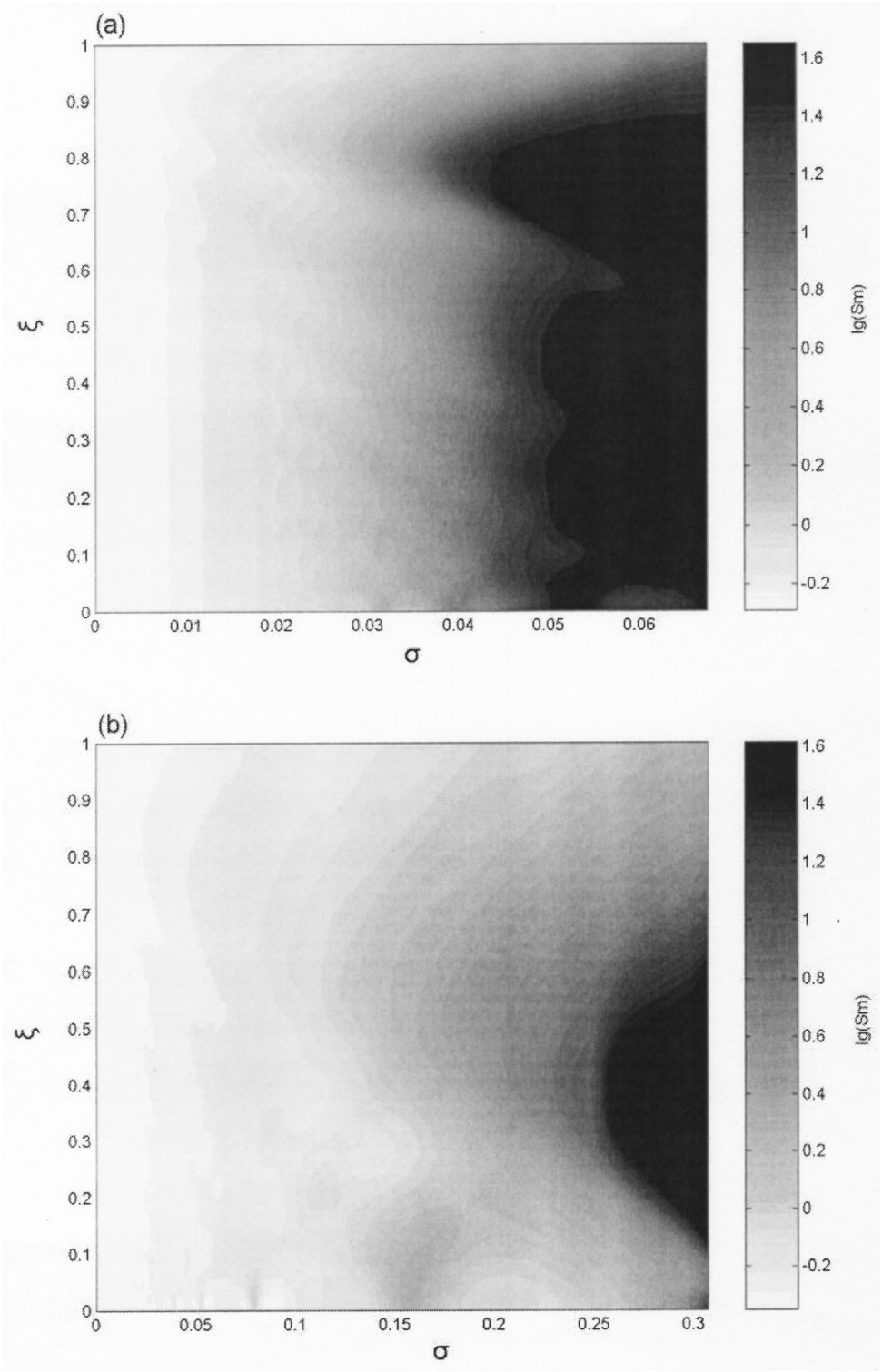


FIG. 7. The spatial distribution of maximum slope S_m. (a) $N=24.083$, $A=0.146\,41$ in region III. (b) $N=3.8533$, $A=0.023\,425$ in region II.

narrow (Fig. 4) and its shock formation distance is similar to that of a spherical wave; we therefore name this one spherical-wave-like shock. Conversely, the waveform in region III is more like that of a plane wave (Fig. 5) just as its shock formation distance approaches that of a plane wave; we name it plane-wave-like shock. Actually, we find that the two kinds of shock can coexist in the piston field just as mentioned in Ref. 23. Figure 6 illustrates the growing process of spherical-wave-like shock and plane-wave-like shock in piston field with parameters $N=6.984$ and $A=0.0425$. It is seen that the plane-wave-like shock forms earlier, and its profile is more like the shock's profile shown in Fig. 5. After the sound beam propagates a certain distance, spherical-wave-like shock appears whose profile is just like the shock's profile shown in Fig. 4. Further propagation of the wave leads the two shocks to move towards, collide, and finally form a sawtooth wave. Before the two shocks collide, they coexist in the sound beam within a long distance. We can see that although spherical-wave-like shock forms later, it can finally grow steeper and narrower than the plane-wave-like

shock since edge wave comes to the axis in a different phase.[23] Therefore, in the near field, the maximum slope is the slope of the plane-wave-like shock, but then, with the development of the spherical-wave-like shock, its slope will exceed the plane-wave-like shock's and the maximum slope is that of the spherical-wave-like shock in the area far from source.

According to the above knowledge of plane-wave-like and spherical-wave-like shocks, we can explain the curve in Fig. 3 as follows: In region III ($r_0 > 6.6 l_D$), because the plane-wave-like shock forms early, its slope exceeds the threshold ϵ first; therefore, with the evaluation of our method, the shock formation distance is the distance where plane-wave-like shock forms, which agrees with a corresponding plane wave's shock formation distance. However, with the decrease of r_0, the influence of diffraction will become evident and the plane-wave-like shock becomes weak, and even in region II ($0.723 l_D < r_0 < 6.6 l_D$), the slope of plane-wave-like shock cannot exceed the threshold at all but the slope of spherical-wave-like shock can still exceed the threshold. So, the shock formation distance in this region is the distance where spherical-wave-like shock forms, which is similar to a spherical wave's shock formation distance. Because the shock formation distances in regions II and III actually represent the shock formation distance of two kinds of shock, there is a break point in the curve of shock formation distance between regions II and III. Finally, in region I ($r_0 < 0.723 l_D$), the influence of diffraction is so strong that both their slopes cannot exceed the threshold; so in this region, the shock formation cannot happen.

Finally, to present the shock formation distance at various radii, two-dimensional spatial distributions of the maximum slope are shown in Fig. 7. It can be seen from Fig. 7(a) ($N = 24.083$, $A = 0.146\,41$, in region III) that the shock formation distances of various radii are almost identical, just like the plane wave. However, Fig. 7(b) ($N = 3.8533$, $A = 0.023\,425$, in region II) shows in some regions the maximum slope at off-axis is larger than that at the axis. It approximately has a spherical distribution near the axis, just as the sound field is radiated from a spherical source. These distributions of maximum slope confirm again that the plane-wave-like shock is similar to a plane wave shock, but the spherical-wave-like shock is similar to a spherical wave shock.

IV. CONCLUSION

In summary, this paper evaluates the shock formation distance of a piston field and concentration is primarily on the diffraction. All simulation results show that the diffraction plays a major role in the shock formation. When the diffraction is small (about within $r_0 > 6.6 l_D$), the shock formation distance of a piston field approaches that of a plane wave; when the diffraction is large (about within $0.723 l_D < r_0 < 6.6 l_D$), the influence of the Rayleigh distance on the shock formation distance is just like that of the spherical source radius on the shock formation distance of a spherical field. Formula (7) is presented to fit this relation. Further, when the diffraction is large enough (about within $r_0 < 0.723 l_D$), the shock cannot even come into being at all. Therefore, the shock formation distance of a piston field can be estimated effectively by that of its corresponding plane wave only when the influence of the diffraction is much smaller than that of the nonlinear intensity.

For giving a reasonable explanation of these phenomena, we study and compare the properties of the shock in sound beam from three aspects, the shock formation distance, the pressure waveform, and the spatial distribution of the S_m value. We find that two kinds of shock coexist in the piston field. One is named plane-wave-like shock, for it is similar to the shock of a plane wave. Another is named spherical-wave-like shock, for it is more like the shock of a spherical wave. Two shocks compete in the piston field. When the plane-wave-like shock is stronger than the spherical-wave-like, the shock formation of the piston field is just like a plane wave. Conversely, it is more like a spherical wave. Although our discussion is based on the piston field in water, the discussion and conclusion can also be similarly made on the other bounded beam with finite amplitude. Because the proposed method of evaluating the shock formation distance considers fully the nonlinear distortion of the waveform, the changes in the wave amplitude, and so on, it is also suitable for the other field, focused or unfocused.

ACKNOWLEDGMENT

This work was supported by NSF of China (Grant Nos. 10074035 and 19834040).

[1] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Plenum, New York, 1977).
[2] S. I. Aanonsen, T. Barkve, J. Naze Tjotta, and S. Tjotta, "Distortion and harmonic generation in the nearfield of a finite amplitude sound beam," J. Acoust. Soc. Am. **75**, 749–768 (1984).
[3] M. F. Hamilton, J. Naze Tjotta, and S. Tjotta, "Nonlinear effects in the farfield of a directive sound source," J. Acoust. Soc. Am. **78**, 202–216 (1985).
[4] T. Kamakura, N. Hamada, K. Aoki, and Yu. Kumamoto, "Nonlinear generated spectral components in the nearfield of a directive sound source," J. Acoust. Soc. Am. **85**, 2331–2337 (1989).
[5] R. O. Cleveland, J. P. Chambers, H. E. Bass, R. Raspet, D. T. Blackstock, and M. F. Hamilton, "Comparison of computer codes for the propagation of sonic boom waveforms through isothermal atmospheres," J. Acoust. Soc. Am. **100**, 3017–3027 (1996).
[6] Y. S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[7] M. A. Averkiou and M. F. Hamilton, "Nonlinear distortion of short pulses radiated by plane and focused circular pistons," J. Acoust. Soc. Am. **102**, 2539–2548 (1997).
[8] J. Tavakkoli, D. Cathignol, R. Souchon, and O. A. Sapozhnikov, "Modeling of pulsed finite-amplitude focused sound beams in time domain," J. Acoust. Soc. Am. **104**, 2061–2072 (1998).
[9] P. T. Christopher and K. J. Parker, "New approaches to nonlinear diffractive field propagation," J. Acoust. Soc. Am. **90**, 488–499 (1991).
[10] F. P. Curra, P. D. Mourad, V. A. Khokhlova, and L. A. Crum, "High intensity focussed ultrasound and tissue heating: the effect of nonlinear sound propagation and vessel presence," Proceedings of IEEE International Ultrasonics Symposium, Sendai, Japan, 1998, pp. 1419–1422.
[11] J. Naze Tjotta, S. Tjotta, and E. H. Vefring, "Propagation and interaction of two collimated finite amplitude sound beams," J. Acoust. Soc. Am. **88**, 2859–2870 (1990).
[12] T. Christopher, "A nonlinear plane-wave algorithm for diffractive propagation involving shock waves," J. Comput. Acoust. **1**, 371–393 (1993).
[13] T. S. Hart and M. F. Hamilton, "Nonlinear effects in focused sound beams," J. Acoust. Soc. Am. **84**, 1488–1496 (1988).
[14] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, San Diego, 1998).

[15] B. Ystad and J. Bernsten, ''Numerical solution of the KZK equation for focusing sources,'' Acta Acust. (Beijing) **3**, 323–330 (1995).
[16] V. A. Khokhlova, M. A. Averkiou, S. J. Yonghouse, M. F. Hamilton, and L. A. Crum, ''Fast Spectral Algorithm for Modeling of Focused Sound Beams in a Highly Nonlinear Regime,'' Proceedings of 16th International Congress on Acoustics and 135th Meeting of Acoustical Society of America, Seattle, WA, 20–26 June 1998, Vol. 4, pp. 2875–2876.
[17] Y. A. Pishchal'nikov, O. A. Sapozhnikov, and V. A. Khokhlova, ''A modification of the spectral description of nonlinear acoustic waves with discontinuities,'' Acoust. Phys. **42**, 362–367 (1996).
[18] V. G. Andreev, Y. A. Pischal'nikov, O. A. Sapozhnikov, V. A. Khokhlova, and R. O. Cleveland, ''Diagnostics of a relaxing medium by an acoustic pulse with shock front,'' Acoust. Phys. **45**, 8–13 (1999).
[19] S. S. Kasheeva, V. A. Khokhlova, O. A. Sapozhnikov, M. A. Averkiou, and L. A. Crum, ''Nonlinear distortion and absorption of intense acoustic waves in media with power frequency law of absorption,'' Acoust. Phys. **46**, 170–177 (2000).
[20] A. C. Baker, A. M. Berg, A. Sahin, and J. Naze Tjotta, ''The nonlinear pressure field of plane rectangular apertures: Experimental and theoretical results,'' J. Acoust. Soc. Am. **97**, 3510–3517 (1995).
[21] T. Kamakura, M. Tani, Y. Kumamoto, and K. Ueda, ''Harmonic generation in finite amplitude sound beams form a rectangular aperture source,'' J. Acoust. Soc. Am. **91**, 3144–3151 (1992).
[22] S. Nachef, D. Cathignol, J. N. Tjotta, A. M. Berg, and S. Tjotta, ''Investigation of a high intensity sound beam from a plane transducer. Experimental and theoretical results,'' J. Acoust. Soc. Am. **98**, 2303–2323 (1995).
[23] V. A. Khokhlova, R. Souchon, J. Tavakkoli, O. A. Sapozhnikov, and D. Cathignol, ''Numerical modeling of finite-amplitude sound beams: Shock formation in the near field of a cw plane piston source,'' J. Acoust. Soc. Am. **110**, 95–108 (2001).
[24] E. A. Zabolotskaya and R. V. Khokhlov, ''Quasi-plane waves in the nonlinear acoustics of confined beams,'' Sov. Phys. Acoust. **15**, 35–40 (1969).
[25] V. P. Kuznetsov, ''Equations of nonlinear acoustics,'' Sov. Phys. Acoust. **16**, 467–470 (1971).

Ray stability in weakly range-dependent sound channels

F. J. Beron-Vera[a] and M. G. Brown
RSMAS/AMP, University of Miami, Miami, Florida 33149

(Received 30 August 2002; accepted for publication 21 April 2003)

Ray stability is investigated in environments consisting of a range-independent background sound-speed profile on which a range-dependent perturbation is superimposed. Theoretical arguments suggest and numerical results confirm that ray stability is strongly influenced by the background sound speed profile. Ray instability is shown to increase with increasing magnitude of $\alpha(I)=(I/\omega)d\omega/dI$, where $2\pi/\omega(I)$ is the range of a ray double loop and I is the ray action variable. This behavior is illustrated using internal-wave-induced scattering in deep ocean environments and rough surface scattering in upward refracting environments. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582444]

PACS numbers: 43.30.Cq, 43.30.Ft, 43.30.Pc [RAS]

I. INTRODUCTION

Measurements made during the Slice89 propagation experiment,[1] made in the eastern North Pacific, showed a clear contrast between highly structured steep ray arrivals and relatively unstructured near-axial flat ray arrivals. Measurements made during the AET experiment[2,3] in a similar environment provided further evidence of the same behavior. Motivated by these observations, several authors[4–6] have investigated ray sensitivity to environmental parameters. The work described in this paper continues the same line of investigation.

Like the earlier work we focus on ray path stability in physical space or phase space. The extension to travel time stability is not considered in this paper. Clearly, however, travel time stability must be addressed to fully understand the Slice89 and AET measurements. In spite of this limitation, and limitations of the ray approximation, our results are in agreement—qualitatively, at least—with the Slice89 and AET observations.

In Sec. II we provide the theoretical background for the work that follows. Most of this material builds on the action-angle form of the ray equations. A trivial observation that follows from the action-angle formalism is that ray sensitivity to the background sound speed structure is controlled by the function $\omega(I)$, where $2\pi/\omega$ is the range of a ray double loop and I is the action variable. A heuristic argument suggests that $d\omega/dI$ should be closely linked to ray stability. In Sec. III we present numerical simulations that are chosen to demonstrate the importance of the background sound speed structure on ray stability. Simulations are shown for both internal-wave-induced scattering in deep ocean environments and rough surface scattering in upward-refracting environments. The latter are included to demonstrate the generality of the arguments presented. These simulations strongly suggest that ray stability is controlled by the magnitude of the nondimensional quantity

$$\alpha(I)=\frac{I}{\omega}\frac{d\omega}{dI}. \tag{1}$$

In Sec. IV we explain the mechanism by which $|\alpha|$ controls ray stability. In Sec. V we summarize our results and briefly discuss: (i) the relationship between our work and earlier investigations; (ii) timefront stability; (iii) the dynamical systems viewpoint; and (iv) the extension to background sound speed structures with range dependence. An appendix is reserved for some mathematical details.

II. THEORETICAL BACKGROUND

A. One-way ray equations

We consider underwater sound propagation in a two-dimensional waveguide with Cartesian coordinates z (upward) and r (along-waveguide). One-way ray trajectories satisfy the *canonical Hamilton's equations* (see, e.g., Ref. 7, and references therein),

$$\frac{dp}{dr}=-\frac{\partial h}{\partial z},\quad \frac{dz}{dr}=\frac{\partial h}{\partial p}, \tag{2}$$

with Hamiltonian

$$h(p,z;r)=-\sqrt{c^{-2}-p^2}.$$

Here, $c(z;r)$ is the sound speed; p is the vertical slowness which is understood as the momentum, conjugate to the generalized coordinate z; and r is the independent (time-like) variable. The vertical slowness and the sound speed are related through $pc=\sin\varphi$, where φ is the angle that a ray makes with the horizontal.

B. Near-integrability under small waveguide perturbations

Assume that the sound speed can be split into a background (range-independent) part, $C(z)$, and a small range-dependent perturbation, $\delta c(z;r)$. Then, to lowest order in $\delta c/c$, the Hamiltonian takes the additive form

$$h=H(p,z)+\delta h(p,z;r).$$

Introduce now the Poincaré action

$$I=\frac{1}{2\pi}\oint dz\,p=\frac{1}{\pi}\int_{z_-}^{z_+}dz\sqrt{C^{-2}-H^2}, \tag{3}$$

[a]Author to whom correspondence should be addressed; electronic mail: fberon@rsmas.miami.edu

where $z_\pm$ denotes the vertical coordinate of the upper $(+)$ and lower $(-)$ turning points of a ray double loop, and consider the canonical transformation into *action-angle variables* $(p,z)\mapsto(I,\vartheta)$ defined by

$$p=\frac{\partial G}{\partial z},\quad \vartheta=\frac{\partial G}{\partial I},\quad G(z,I)=\int_{z_-}^{z} d\xi\sqrt{C^{-2}(\xi)-H^2}.$$

According to the above-given transformation,

$$h(p,z;r)\mapsto\overline{H}(I)+\delta\overline{h}(I,\vartheta;r)$$

and the ray equations (2) take the form

$$\frac{dI}{dr}=-\frac{\partial}{\partial\vartheta}\delta\overline{h},\quad \frac{d\vartheta}{dr}=\omega+\frac{\partial}{\partial I}\delta\overline{h}, \tag{4}$$

where

$$\omega(I)=\frac{d\overline{H}}{dI}. \tag{5}$$

Set (4) constitutes a *near-integrable Hamiltonian system* for the reasons explained next.

In the limit $\delta\overline{h}\to 0$, the ray equations (4), which have one degree of freedom, are autonomous and the corresponding Hamiltonian, $\overline{H}$, is an integral of motion that constrains the dynamics. As a consequence, the equations are integrable through quadratures and the motion is *periodic* with (spatial) frequency ω. Namely $I=I_0$ and $\vartheta=\vartheta_0+\omega r \mod 2\pi$, where I_0 and ϑ_0 are constants. Every solution curve is thus a line that winds around an invariant one-dimensional torus, whose representation in (p,z) space is the closed curve given by the isoline $H=\overline{H}(I_0)$. Notice that $C^{-1}\cos\varphi=H$; consequently, each torus can be uniquely identified by the ray axial angle defined by $\varphi_a(p,z)=\arccos C_a H$, where C_a is the background sound speed at the sound channel axis.

With nonzero $\delta\overline{h}$, the Hamiltonian, $\overline{H}+\delta\overline{h}$, is no longer an invariant (the equations are nonautonomous) and the system may be *sensitive to initial conditions*, leading to *chaotic motion* in phase space. The distinction between regular and chaotic ray trajectories is commonly quantified by the *Lyapunov exponent*, formally defined by

$$\nu_\infty=\lim_{r\to\infty}\lim_{d_0\to 0}\frac{1}{r}\ln\frac{d}{d_0}, \tag{6}$$

where $d(r)$, such that $d(0)=d_0$, is a suitably chosen measure of the separation between neighboring ray trajectories in phase space. For regular trajectories, $d\sim r^a$ as $r\to\infty$ with $a>0$ and, hence, $\nu_\infty=0$. For chaotic trajectories, instead, $d\sim\exp\nu_\infty r$; in this case, ν_∞^{-1}, the average e-folding range, is regarded as the *predictability horizon*.

C. Variational equations

Although any norm of $(\delta p,\delta z)$, a perturbation to a trajectory (p,z), could be used to define a distance in phase space, this is not a trivial task because z and p do not have the same dimensions, which complicates the evaluation of Eq. (6). The following procedure eliminates this problem.

The *variational equations*, which follow from the ray equations (2), are

$$\frac{d\mathbf{Q}}{dr}=\mathbf{J}\mathbf{Q},\quad \mathbf{Q}(0)=\mathbf{I}, \tag{7}$$

where

$$\mathbf{J}=\begin{bmatrix}-\dfrac{\partial^2 h}{\partial p\,\partial z} & -\dfrac{\partial^2 h}{\partial z^2}\\[2ex] \dfrac{\partial^2 h}{\partial p^2} & \dfrac{\partial^2 h}{\partial p\,\partial z}\end{bmatrix},\quad \mathbf{Q}=\begin{bmatrix}\dfrac{\partial p}{\partial p_0} & \dfrac{\partial p}{\partial z_0}\\[2ex] \dfrac{\partial z}{\partial p_0} & \dfrac{\partial z}{\partial z_0}\end{bmatrix},$$

with $p_0=p(0)$ and $z_0=z(0)$, and $\mathbf{I}$ is the identity matrix. Here, $\mathbf{J}(r)$ and $\mathbf{Q}(r)$ are the Jacobian matrices of the Hamiltonian vector field and associated flow, respectively; the last is usually referred to as the *stability matrix*. Notice that $(\delta p,\delta z)^T=\mathbf{Q}(\delta p_0,\delta z_0)^T$ at the lowest order in $(\delta p_0,\delta z_0)$. Let now $\nu^{\mathbf{Q}}(r)$ be the largest of the two eigenvalues of $\mathbf{Q}$, and consider the definition (see, e.g., Ref. 8)

$$\nu_\infty=\lim_{r\to\infty}\frac{1}{r}\ln|\nu^{\mathbf{Q}}|. \tag{8}$$

If the limit in Eq. (8) exists, and is not nil, then $(\delta p,\delta z)\sim\exp\nu_\infty r$ as $r\to\infty$ (nearby trajectories diverge exponentially in range) and, hence, Eq. (8) can be taken as a suitable definition of the Lyapunov exponent. The equivalence between Eqs. (6) and (8) can be understood by noting that the variational equations describe the evolution of an infinitesimal circle of initial conditions surrounding a specified ray initial condition. The circle gets deformed into an ellipse whose area equals that of the initial circle according to Liouville's theorem (see, e.g., Refs. 9 and 10). The eigenvectors of $\mathbf{Q}$ define the orientation of the ellipse. The largest eigenvalue is a measure of the length of the semimajor axis and, hence, is a suitable choice of d.

A simple but very important observation follows from the action-angle formalism. Dependence of the ray and variational equations on the background sound speed structure enters only through the function $\omega(I)$. The action-angle form of the variational equations for the perturbed system strongly suggests that ray stability and $d\omega/dI$ are closely linked. The mechanism through which $d\omega/dI$ influences ray stability can be seen from the action-angle form of the ray variational equations,

$$\frac{d}{dr}\delta I=-\frac{\partial^2\delta\overline{h}}{\partial I\,\partial\vartheta}\delta I-\frac{\partial^2\delta\overline{h}}{\partial\vartheta^2}\delta\vartheta,$$

$$\frac{d}{dr}\delta\vartheta=\frac{d\omega}{dI}\delta I+\frac{\partial^2\delta\overline{h}}{\partial I^2}\delta I+\frac{\partial^2\delta\overline{h}}{\partial I\,\partial\vartheta}\delta\vartheta.$$

If one assumes that the second derivatives of $\delta\overline{h}$ are zero-mean random variables, then when $d\omega/dI=0$ these terms should lead to slow (power-law) growth of $\delta\vartheta$ and δI. But if $|d\omega/dI|$ is large, this term will cause $|\delta\vartheta|$ to rapidly grow for any nonzero $|\delta I|$. The perturbation terms will then lead to a mixing of $|\delta\vartheta|$ and $|\delta I|$. The term $d\omega/dI$ will lead, in turn, to further growth of $|\delta\vartheta|$. As this process repeats itself, both $|\delta I|$ and $|\delta\vartheta|$ are expected to grow rapidly. Thus ray instability is expected to be significantly enhanced when $|d\omega/dI|$ is large.

D. Kolmogorov–Arnold–Moser theory

A (background) Hamiltonian is said to be *nondegenerate* if $d\omega/dI \neq 0$, i.e., if the frequency varies from torus to torus, which is a condition for nonlinearity of the system. An important result for near-integrable nondegenerate Hamiltonian systems is the celebrated *Kolmogorov–Arnold–Moser* (KAM) *theorem* on the stability of periodic solutions; see, e.g., Refs. 9 and 10. This theorem states that if $\delta\overline{h}$ is small enough, for most initial conditions the motion remains periodic (i.e., confined to tori) and the complement of the periodic motion (i.e., the chaotic motion) has a measure that tends to zero as $\delta\overline{h} \to 0$. The KAM theorem thus guarantees that periodic motion (i.e., the KAM tori) separate the destroyed tori, leading to the notion of ''islands'' of (eternal) stability immersed in a chaotic ''sea.''

The mechanism that produces the destruction of tori is trajectory-medium resonance; see, e.g., Refs. 9 and 10. For example, assume the perturbation Hamiltonian $\delta\overline{h}$ to be periodic in range with frequency Ω. Then it can be represented in Fourier series $\delta\overline{h} = \mathrm{Re}\,\Sigma_{m,n} A_{mn}(I) \exp i(m\vartheta - n\Omega r)$. For KAM or *nonresonant* tori, $m\omega + n\Omega \neq 0$ for all integers n,m and the motion is periodic. In contrast, those tori that satisfy $m\omega + n\Omega = 0$ for some integers n,m are said to be in *resonance* and chaotic motion develops. If several tori are captured into resonance, then the character of the chaotic motion will depend on whether these resonances overlap or not. For instance, consider two resonances centered at I_1 and I_2. The widths of these resonances can be estimated as $\Delta I_i = 4\sqrt{A_i/|d\omega_i/dI|}$, where A_i is the amplitude of the resonant term in the above-given expansion. Define then $\Delta I = \Delta I_1 + \Delta I_2$. A benign form of chaos is present when these resonances are isolated, i.e., $\Delta I < |I_1 - I_2|$; strong chaos, in contrast, emerges when these resonances overlap, i.e., $\Delta I > |I_1 - I_2|$. The last criterion, due to Chirikov,[11] thus gives a quantitative estimate of the size of the region of phase space occupied by chaotic trajectories.

The focus in KAM theory on the role of individual ray-medium resonances might seem to be at odds conceptually with the heuristic argument given at the end of Sec. II C. There it was argued that the second derivatives of the perturbation to the environment could be treated as random variables. Reference 12 partially bridges this conceptual gap by showing that the KAM theorem can be applied to problems for which the perturbation consists of a superposition of an arbitrarily large finite number of frequencies. Chirikov's criterion is still applicable, but its evaluation seems feasible only if the number of frequencies that comprises the perturbation is very small.

III. NUMERICAL SIMULATIONS

In the following, numerical simulations are presented which were chosen to demonstrate the importance of the background sound speed structure on ray stability. We consider first deep-water conditions and upward-refracting conditions afterward.

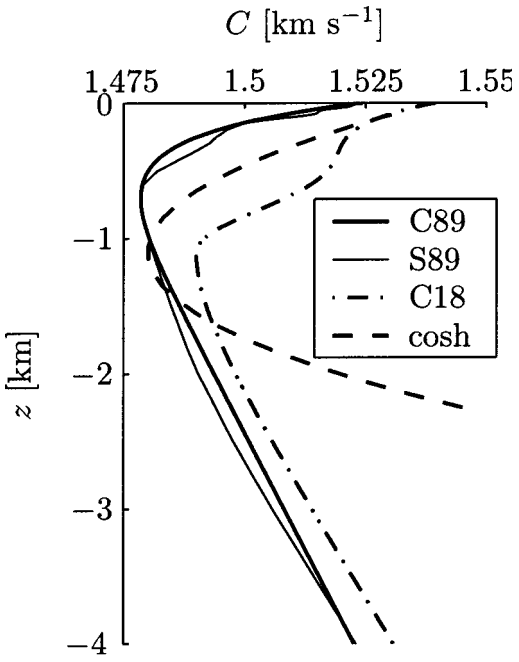


FIG. 1. Background sound speed profiles used in the numerical work presented in this paper.

A. Deep-water conditions

Four different background profiles are studied here (cf. Fig. 1). These are designated S89, C89, C18, and cosh. The S89 profile corresponds to a range average of the Slice89 sound speed observations. The C89 profile is a canonical profile[13] with parameters chosen to approximately match S89's sound channel axis depth, axial sound speed, and surface sound speed. The C18 profile can be regarded as an idealized model of the sound speed structure in the North Atlantic, whose upper ocean structure is associated with the 18 °C water mass.[7] Finally, the cosh profile, with cosh-dependence on depth relative to the axis, was chosen because it has the special property $d\omega/dI = 0$ for all I.

The same internal-wave-induced sound speed perturbation field is superimposed on all four background profiles. This field is assumed to satisfy the relationship

$$\delta c/c = \mu N^2 \zeta, \tag{9}$$

where $\zeta(z;r)$ is the internal-wave-induced vertical displacement of a water parcel and $N(z;r)$ is the Brunt–Väisälä frequency. Relationship (9) with $\mu = 1.25\ \mathrm{s^2\,m^{-1}}$ was found to give a good fit to AET hydrographic measurements.[14] Our simulated internal-wave-induced sound speed perturbations are similar to those used in Refs. 7 and 15; these are based on Eq. (9) and make use of the N profile estimated from measurements made during the AET experiment. The statistics of ζ are assumed to be described by the empirical Garrett–Munk spectrum.[16] The vertical displacement ζ is computed using Eq. (19) of Ref. 17 with the variable x replaced by r and $y = 0 = t$. Physically this corresponds to a frozen vertical slice of the internal wave field that includes the influence of transversely propagating internal wave modes. A Fourier method is used to numerically generate the sound speed perturbation fields. A mode number cutoff of 30 and a horizontal

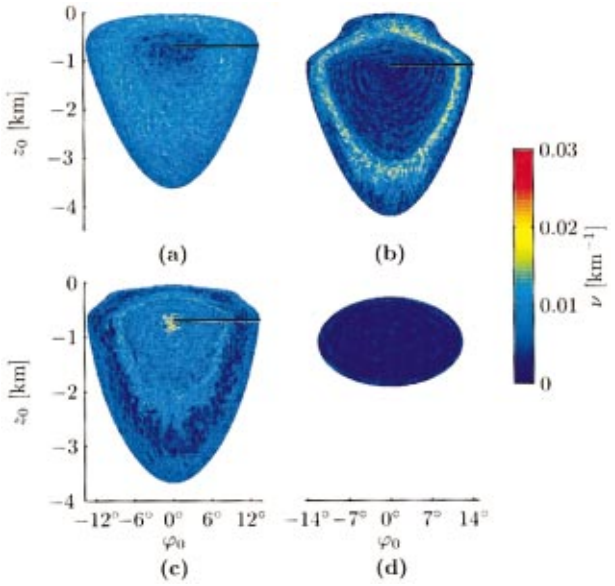


FIG. 2. Stability exponents (finite range estimates of Lyapunov exponents) as a function of the initial depth and launch angle, for ray simulations in the (a) C89, (b) C18, (c) S89, and (d) cosh waveguides (see Fig. 1) with internal-wave-induced perturbations superimposed.

wave number cutoff of 2π km^{-1} were used in our simulations. The internal wave strength parameter was taken to be the nominal Garrett–Munk value.

Figure 2 shows the stability exponent, ν (a finite-range estimate of the Lyapunov exponent; cf. the appendix), as a function of initial ray position in phase space for each of the four environments considered. Figure 2 provides a general picture of the ray motion stability character in each of the waveguides. In the C89 waveguide, ray trajectories with small (respectively, large) unperturbed ray axial angles have small (respectively, large) associated stability exponents. This trend is reversed in the S89 waveguide. The disparity in the stability properties of the ray motion in these waveguides contrasts with the close similarity of the corresponding background sound speed profiles. In the C18 waveguide ray trajectories have in general relatively small associated stability exponents, except in a narrow band of initial actions (or unperturbed axial angles) where the exponents are large and the ray motion more unstable. In opposition to the other waveguides, in the cosh profile the stability exponents are very small everywhere in phase space and the ray motion is remarkably stable.

In each panel of Fig. 2 the horizontal line shown corresponds to a fan of rays, launched on the sound channel axis, with positive angles. For these rays, in each environment ray depth at a range of 1000 km and stability exponent are plotted as a function of launch angle, φ_0, in Fig. 3. In Fig. 3 bands of regular trajectories appear as a smooth slowly varying curve $z(\varphi_0)$ with small associated values of ν; bands of chaotic rays appear as a highly structured $z(\varphi_0)$ curve with large associated values of ν. Also shown in Fig. 3 is a plot of the stability parameter α defined in Eq. (1) vs φ_0 in each environment. (The action I is a monotonically increasing function of the axial ray angle which, for rays starting at the axis, coincides with φ_0; thus replacing I by φ_0 represents a simple stretching of the abscissa.) Notice in the z vs φ_0 plots the seemingly unstructured (respectively, ordered) distribution of points associated with those angular bands where ν is large (respectively, small). Notice also that ν (or, equivalently, the irregularity of z as a function of φ_0) appears to increase with increasing $|\alpha|$. In particular, note that $\alpha=0$ (since $d\omega/dI=0$) for all φ_0 in the cosh profile. In this case ray final depth varies very slowly with φ_0, and the stability exponents are very small.

The numerical results presented in Figs. 2 and 3 suggest that ray stability is strongly influenced by the background sound speed structure and that ray instability increases with increasing $|\alpha|$. These results are consistent with the heuristic argument given in Sec. II based on the action-angle form of the ray variational equations; there it was argued that ray instability should increase with increasing $|d\omega/dI|$.

B. Upward-refracting conditions

The validity of the argument given in Sec. II is not limited to deep ocean conditions, strongly suggesting that the result should be more generally valid. We now present numerical results that support this expectation.

Figure 4 shows two upward-refracting sound speed profiles. Figure 5 shows α in the same environments and the difference in range, Δr, between perturbed (rough surface) and unperturbed (flat surface) rays as a function of launch angle at the surface after making 21 loops, which corresponds to 20 surface reflections. [In this type of environment the definition of I in Eq. (3) is unchanged except that the upper integration limit is $z=0$ for all rays.] Rough surface scattering was treated using a frozen simulated surface gravity wave field with a $k^{-7/2}$ surface elevation wave number spectrum with $0.02\ \text{rad m}^{-1}\leqslant k\leqslant 0.16\ \text{rad m}^{-1}$, $\Delta k = 10^{-3}\ \text{rad m}^{-1}$, and rms slope of 4×10^{-3}. To treat specular ray reflections from this surface, the surface boundary condition was linearized; the surface elevation was neglected, but the nonzero slope was not approximated.

Figure 5 shows clearly that ray stability, as measured by Δr, is controlled almost entirely by the background sound speed structure via α, rather than details of the rough surface.

IV. SHEAR-INDUCED RAY INSTABILITY

In this section we present additional numerical simulations that give insight into the mechanism by which α influences ray instability. For simplicity we assume deep water conditions here. The arguments presented here make use of the well-known (see, e.g., Refs. 9 and 10) analogy between ray motion, defined by Eq. (2) and particle motion in an incompressible two-dimensional fluid.

The left panels of Fig. 6 show the range evolution of a segment of a *Lagrangian manifold* (a smooth curve in phase space) in the unperturbed C89 (top) and cosh (bottom) waveguides. The segment is depicted with a dashed line at $r=0$ and with a solid line at $r=1000$ km in both waveguides. As a consequence of Liouville's theorem the segment cannot break or intersect itself but it can increase in complexity as range increases. In the unperturbed case, since the motion is integrable (i.e., each point of the segment preserves its initial I), the length of the segment can grow in

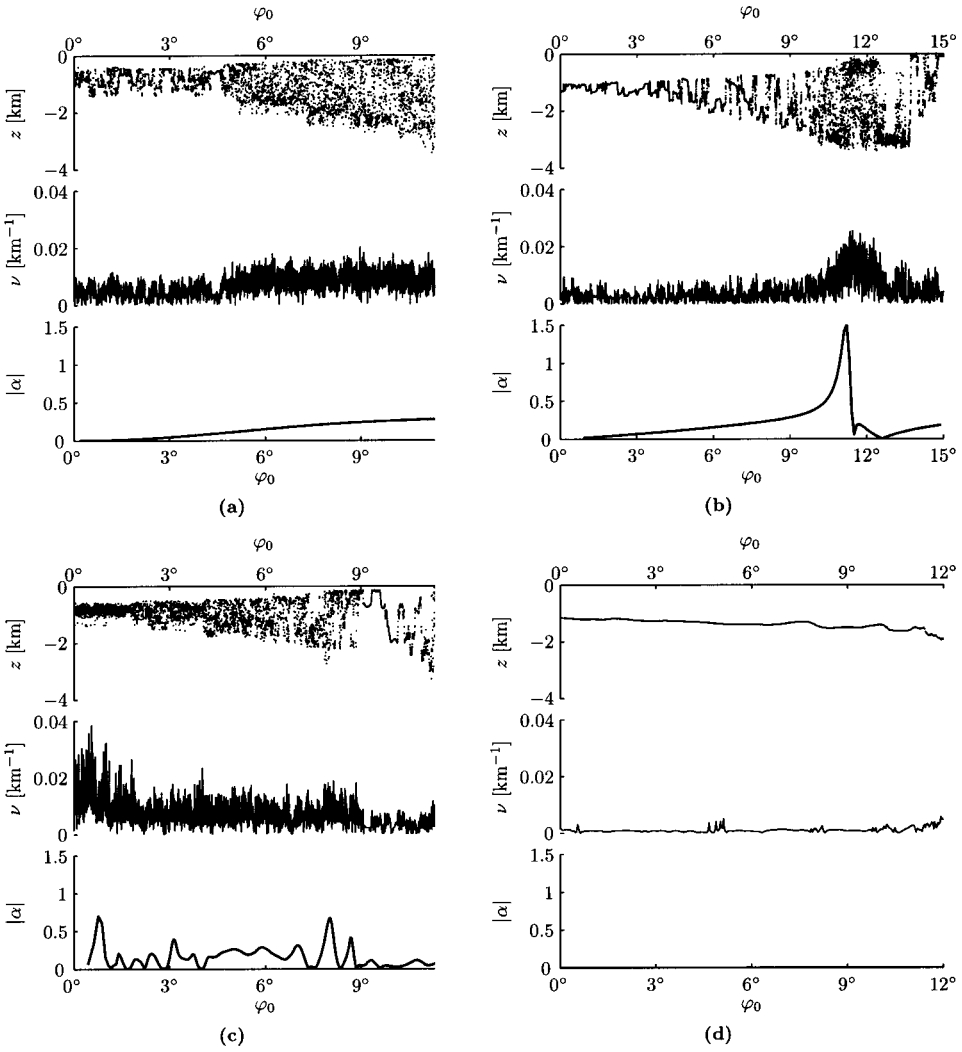


FIG. 3. Each panel shows ray final depth (upper plot), stability exponent (middle plot), and stability parameter (lower plot) at 1000 km range as a function of ray launch angle for a source on the sound channel axis. The sound speed structures are identical to those used in the panels of Fig. 2 with the same labels.

range, at most, following a power law. The cosh profile has a special property. In that profile the manifold segment just rotates counterclockwise at a constant rate ω. The reason for the difference in behavior is that in the C89 waveguide ω varies with I, whereas in the cosh profile ω does not vary with I. The monotonic decay of ω as a function of I in the C89 waveguide induces a ''shear'' in phase space which causes the outermost points of the segment to rotate more slowly than the innermost ones and, hence, causes the segment to spiral. The ray motion in phase space associated with the unperturbed C89 waveguide can thus be regarded as analogous to that of ideal fluid particles passively advected by a stationary planar circular flow with radial shear. In the cosh profile there is no shear. In polar coordinates *radial shear* can be defined as

$$\rho \frac{\partial}{\partial \rho}\left(\frac{u_\theta}{\rho}\right), \tag{10}$$

where ρ is the radial coordinate and u_θ is the θ component of the velocity field. (More correctly, this quantity is, apart from a factor of 2, the $\rho\theta$ component of the strain-rate tensor for planar circular flow; see, e.g., Ref. 18.) The connection with ray motion in phase space can be accomplished by identifying I with ρ and ωI with u_θ. The replacements $\rho \mapsto I$ and $u_\theta \mapsto \omega I$ in Eq. (10) thus give the analogous expression $I d\omega/dI$ for the shear in phase space. Notice that this expression is (apart from the ω^{-1} factor) the stability parameter α. We have chosen to include the ω^{-1} factor in the definition of α because of precedent[19] and because it is convenient to make α dimensionless.

The right panels of Fig. 6 show the evolution of the Lagrangian manifold segment in the same waveguides as those used to produce the left panels but with a superimposed perturbation induced by internal wave fluctuations. Notice the highly complicated structure of the Lagrangian manifold in the perturbed C89 waveguide as compared to that in the unperturbed one. (Note that the fan of rays used to produce the figure is far too sparse to resolve what should be an unbroken smooth curve which does not intersect itself.) In contrast, observe that in the cosh environment the sound speed perturbation has only a very minor effect on the evolution of the Lagrangian manifold.

Perturbations to steeper rays caused by internal-wave-induced sound speed perturbations in deep ocean environments, including those in our simulations, are significant only near the ray's upper turning depth. This observation motivates a simple model that gives insight into the mechanism by which α is linked to ray stability. In the model, each portion of a Lagrangian manifold acquires a sinusoidal ''wrinkle'' at each upper turning point, but is otherwise un-

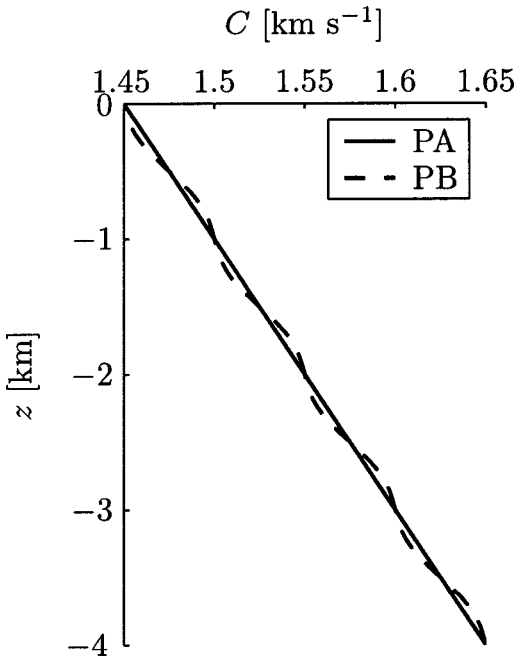


FIG. 4. Background sound speed profiles used to construct the curves in Fig. 5.

affected by the sound speed perturbations. The evolution of small segments of a Lagrangian manifold using such a model, in both the C89 and cosh waveguides, is shown in Fig. 7. Notice the rapid growth in complexity of the segment as the range increases in the C89 waveguide. After acquiring a wrinkle, the segment stretches and folds as a result of the radial shear in phase space. As additional wrinkles are acquired this process is repeated successively in range, making the shape of the Lagrangian manifold segment even more complex. In opposition to this situation, observe the simplicity of the segment's shape in the cosh waveguide as range increases. In this case, after acquiring a wrinkle, the distorted segment rotates counterclockwise, without the additional influence of shear-induced stretching, at a constant frequency ω.

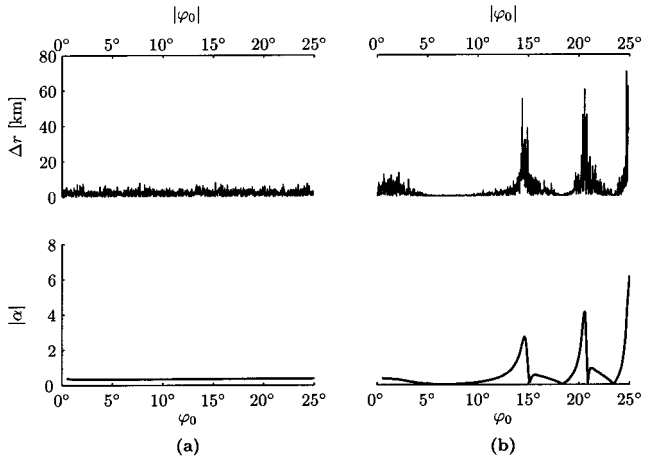


FIG. 5. Upper panels: absolute value of the difference between perturbed (rough surface) and unperturbed (flat surface) ray range after 20 surface reflections as a function of initial ray angle at the surface for sound speed profiles PA (a) and PB (b) in Fig. 4. Lower panels: stability parameter as a function of ray angle in each of the two environments.

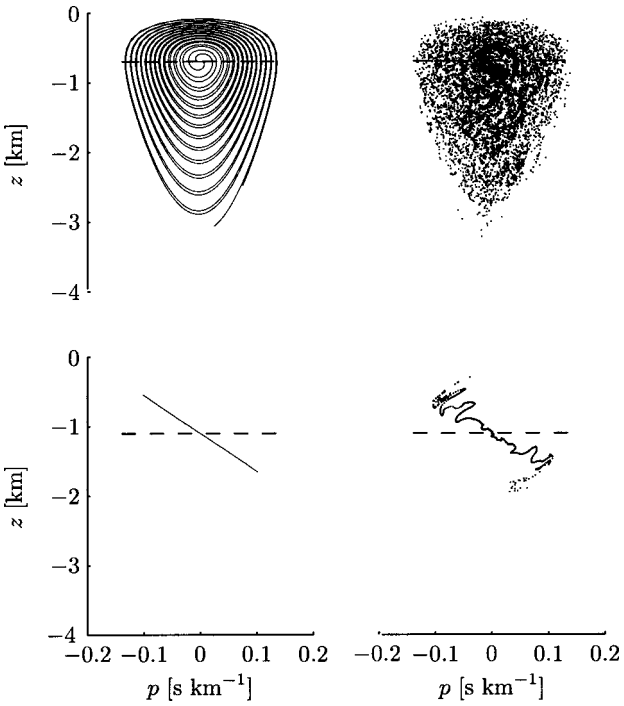


FIG. 6. Evolution of a portion of Lagrangian manifold in the C89 (upper panels) and cosh (lower panels) waveguides with (right panels) and without (left panels) internal-wave-induced perturbations. Dashed and solid/dotted curves show the manifold at $r=0$ and $r=1000$ km, respectively.

Each time a perturbation is introduced, the action I changes by the amount δI, say, which we assume to be of the same order as the perturbation. As a consequence, to lowest order $R(=2\pi/\omega)$ experiences the change

$$R \mapsto (1-\alpha\,\delta I/I)R.$$

The perturbation to the range of a ray double loop R depends on both the perturbation *and* the properties of the background sound speed structure. Under the change $I \mapsto I+\delta I$, a sufficient condition for R to remain invariant at lowest order is $\alpha=0$. This provides an explanation for the remarkable stability of the cosh waveguide. That is, when $\alpha=0$ the ray motion remains periodic at lowest order—no matter the com-

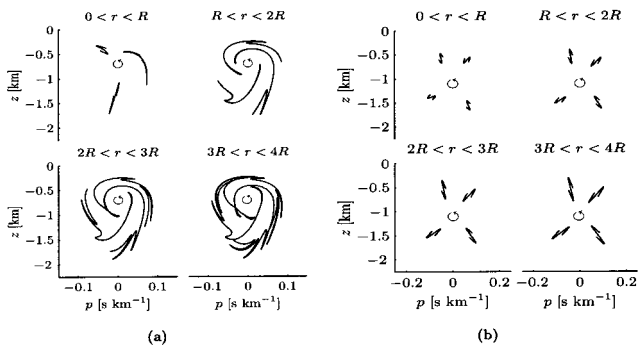


FIG. 7. Evolution of a segment of Lagrangian manifold in the C89 (a) and cosh (b) waveguides under the influence of an idealized range-dependent perturbation. The perturbations are assumed to be in the form of a series of ''kicks'' that produce a sinusoidal deformation to the Lagrangian manifold at each upper turning depth. A portion of Lagrangian manifold, originally located on a torus with frequency $\omega\,(=2\pi/R)$, immediately after experiencing a kick at range $r=(k-1)R$, $k=1,...,4$, and before receiving the next one at $r=kR$, as well as at several intermediate stages, is shown.

plexity of the perturbation mounted on the waveguide. To lowest order, a nonvanishing shear ($\alpha \neq 0$) appears as a necessary condition to sustain the successive stretching and folding of the Lagrangian manifold after it gets distorted by the perturbation. (Of course, chaotic motion is still possible when $\alpha = 0$ provided that the perturbation strength is sufficiently large.) It is thus expected that where $|\alpha|$ is small (respectively, large) there will be less (respectively, more) sensitivity to initial conditions and, hence, the motion be more regular (respectively, chaotic). Support for this conjecture is given in the numerical simulations presented in this paper.

V. SUMMARY AND DISCUSSION

In this paper we have considered ray motion in environments consisting of a range-independent background sound speed profile on which a weak range-dependent perturbation is superimposed. The results presented show that ray stability is strongly influenced by the background sound speed structure; ray instability was shown to increase with increasing magnitude of $\alpha(I) = (I/\omega) d\omega/dI$, where $2\pi/\omega(I)$ is the range of a ray double loop and I is the ray action variable. This conclusion is based largely on numerical simulations, but the simulations were shown to support a simple heuristic argument based on the action-angle form of the ray variational equations. The mechanism by which α controls ray instability was shown to be shear-induced enhancement of perturbations caused by the sound speed perturbation term. The importance of α was illustrated with numerical simulations of ray motion in deep ocean environments including internal-wave-induced scattering, and in upward-refracting environments including rough surface scattering. So far as we are aware, this conclusion is consistent with all of the numerical results presented earlier.[4–6] Reference 6 relied heavily on results that follow from a dynamical systems viewpoint.

The connection between our work and results relating to dynamical systems deserves further comment. Recall that the condition $d\omega/dI \neq 0$ (the nondegeneracy condition) must be satisfied for the KAM theorem to apply, and that this result guarantees that some rays are nonchaotic provided the strength of the range-dependent perturbation is sufficiently weak. This theorem might seem to conflict with our assertion that ray instability increases with increasing $|\alpha|$. This apparent conflict can be resolved by interpreting our statement as a statement of what happens for most rays. That is, for most rays stability exponents (finite range estimates of Lyapunov exponents) increase as $|\alpha|$ increases. This does not rule out the possibility that for a fixed $d\omega/dI \neq 0$ some rays will be nonchaotic.

A seemingly more troublesome conflict between our simulations and KAM theory follows from the result, noted earlier, that each isolated resonance has a width proportional to $|d\omega/dI|^{-1/2}$. Doesn't this imply that rays should become increasingly chaotic as $|d\omega/dI|$ decreases? The answer, we believe, is no. To understand why, consider rays in a band $I_0 \leqslant I \leqslant I_1$ over which $d\omega/dI$ is approximately constant, $d\omega/dI = (\omega_1 - \omega_0)/(I_1 - I_0)$. Within this band resonances are excited at selected values of ω. The number of resonances excited is approximately proportional to $|\omega_1 - \omega_0|$, which, in turn, is proportional to $|d\omega/dI|$. The "degree of chaos" should be proportional to the product of the number of resonances excited and the width of individual resonances. This product scales like $|d\omega/dI|^{1/2}$; this suggests that rays should become increasingly chaotic as $|d\omega/dI|$ increases, consistent with our simulations. Neither the KAM theorem nor the resonance width estimate applies in the limit $d\omega/dI = 0$. Behavior in that limit is likely problem-dependent (e.g., Refs. 20 and 21). Our simulations—which are probably representative of problems characterized by an inhomogeneous background and a weak perturbation with a broad spectrum—suggest that ray motion in this limit is very stable.

In this paper we have addressed the issue of ray stability in physical space or phase space; we have not addressed the related problem of travel time stability. The latter problem is more difficult inasmuch as that problem involves, in addition to the (one-way) ray equations (2), a third equation $dT/dr = L$ and imposition of an eigenray constraint. Here, $L = p\,dz/dr - h(p,z;r)$ using standard variables, or $L = I\,d\vartheta/dr - \overline{h}(I,\vartheta;r)$ using action-angle variables. We have seen some numerical evidence that ray instability in phase space is linked to large time spreads, but this connection is currently not fully understood.

An advantage of our use of the action-angle formalism is that essentially the same results apply if the assumption of a background range-independent sound speed profile, i.e., $C = C(z)$, is relaxed to allow for a slowly varying (in range) background sound speed structure, i.e., $C = C(z;\varepsilon r)$. (Here, ε scales like the ratio of a typical correlation length of the range dependence to a typical ray double loop range.) In the latter case the action is not an exact ray invariant but it is an adiabatic invariant, i.e., $dI/dr = O(\varepsilon^2)$. Thus, correct to $O(\varepsilon)$ the latter problem can be treated as being identical to the former one. Consequently, the problem that we have treated here also applies to slowly varying background environments.

ACKNOWLEDGMENTS

We thank J. Colosi, S. Tomsovic, A. Virovlyansky, M. Wolfson, and G. Zaslavsky for the benefit of discussions on ray chaos. This research was supported by code 321OA of the Office of Naval Research.

APPENDIX: STABILITY EXPONENTS

Liouville's theorem (see, e.g., Refs. 9 and 10) guarantees that areas in the present two-dimensional phase space are preserved, leading to $\det \mathbf{Q} = 1$ as a corollary. Consequently, $2\nu_{\pm}^{\mathbf{Q}} = \mathrm{trace}\,\mathbf{Q} \pm \sqrt{\mathrm{trace}^2\,\mathbf{Q} - 1}$.[22] The condition $\det \mathbf{Q} = 1$, however, is difficult to fulfill in highly chaotic flows due to the limitation of machine numerical precision. More precisely, an initial area in phase space tends to align along the one-dimensional perturbation subspace spanned by the eigenvector of $\mathbf{Q}$ associated with the largest Lyapunov exponent, making the elements of $\mathbf{Q}$ ill-conditioned. This implies that neither $\det \mathbf{Q}$ nor $\mathrm{trace}\,\mathbf{Q}$—and, hence, neither $\nu^{\mathbf{Q}}$ nor ν_∞—can be computed reliably at long range. The common

approach to overcome this problem involves successive normalizations of the elements of **Q**, while integrating simultaneously Eqs. (2) and (7), after a fixed number of range steps.[8] An alternative strategy, which does not involve normalizations, has been taken here.[23]

Let $\mathbf{A}(r)$ be the bounded-element matrix such that

$$\mathbf{Q}=\mathbf{A}e^{\lambda r}, \tag{A1}$$

which implies

$$\nu^{\mathbf{Q}}=\nu^{\mathbf{A}}e^{\lambda r}.$$

Here, λ is a guessed value of ν_∞. The reason for introducing the decomposition (A1) is that if λ is close to ν_∞, the matrix **A** will remain well conditioned at ranges long past those at which **Q** becomes poorly conditioned. This leads, in turn, to significantly improved numerical stability. Equation (7) leads to the *modified variational equations*

$$\frac{d\mathbf{A}}{dr}=(\mathbf{J}-\lambda\mathbf{I})\mathbf{A}, \quad \mathbf{A}(0)=\mathbf{I}. \tag{A2}$$

Notice that $\det\mathbf{Q}=\det\mathbf{A}\exp\lambda r=1$ and, hence, $\det\mathbf{A}\sim 0$ as $r\to\infty$. Equation (A2) has been integrated in this paper after choosing λ, in order to compute a finite range estimate of ν_∞, which we call a *stability exponent* and denote by ν.

[1] T. F. Duda, S. M. Flatté, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, P. F. Worcester, B. M. Howe, J. A. Mercer, and R. C. Spindel, "Measured wavefront fluctuations in 1000 km pulse propagation in the Pacific Ocean," J. Acoust. Soc. Am. **92**, 939–955 (1992).

[2] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, J. A. Colosi, B. M. Howe, J. A. Mercer, A. B. Baggeroer, and K. Metzger, "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the Eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201 (1999).

[3] J. A. Colosi *et al.*, "Comparison of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the Eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218 (1999).

[4] T. F. Duda and J. B. Bowlin, "Ray-acoustic caustic formation and timing effects from ocean sound-speed relative curvature," J. Acoust. Soc. Am. **96**, 1033–1046 (1994).

[5] J. Simmen, S. M. Flatté, and G. Yu-Wang, "Wavefront folding, chaos and diffraction for sound propagation through ocean internal waves," J. Acoust. Soc. Am. **102**, 239–255 (1997).

[6] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, "Theory and application of ray chaos to underwater acoustics," Phys. Rev. E **64**, 1–20 (2001).

[7] M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. Wolfson, and G. M. Zaslavsky, "Ray dynamics in ocean acoustics," J. Acoust. Soc. Am. (in press).

[8] T. S. Parker and L. O. Chua, *Practical Numerical Algorithms for Chaotic Systems* (Springer, New York, 1989).

[9] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1989).

[10] M. Tabor, *Chaos and Integrability in Nonlinear Dynamics* (Wiley, New, York, 1989).

[11] B. V. Chirikov, "A universal instability of many-dimensional oscillator systems," Phys. Rep. **52**, 265–379 (1979).

[12] M. G. Brown, "Phase space structure and fractal trajectories in one-and-a-half degree of freedom Hamiltonian systems whose time dependence is quasiperiodic," Nonlinear Processes Geophys. **5**, 69–74 (1998).

[13] W. H. Munk, "Sound channel in an exponentially stratified ocean with application to SO-FAR," J. Acoust. Soc. Am. **55**, 220–226 (1974).

[14] M. A. Wolfson and J. L. Spiesberger, "Full wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations," J. Acoust. Soc. Am. **106**, 1293–1306 (1999).

[15] F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in a long-range acoustic propagation experiment," J. Acoust. Soc. Am. (submitted).

[16] W. H. Munk, "Internal waves and small scale processes," in *Evolution of Physical Oceanography*, edited by B. Warren and C. Wunsch (MIT, Cambridge, 1981), pp. 264–291.

[17] J. A. Colosi and M. G. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound speed perturbation fields," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).

[18] G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge University Press, New York, 1964).

[19] G. M. Zaslavsky, *Physics of Chaos in Hamiltonian Systems* (Imperial College, London, 1998).

[20] A. A. Chernikov, M. Y. Natenzon, B. A. Petrovichev, R. Z. Sagdeev, and G. M. Zaslavsky, "Some peculiarities of stochastic layer and stochastic web formation," Phys. Lett. A **122**, 39–46 (1987).

[21] A. Chernikov, R. Sagdeev, and G. Zaslavsky, "Chaos: How regular can it be?," Phys. Today **41**, 27–35 (1988).

[22] It can be shown that $\nu_{\pm}^{\mathbf{Q}}\sim\mathrm{trace}^{\pm 1}\mathbf{Q}$ as $r\to\infty$. Accordingly, Eq. (8) can be replaced by the equivalent expression $\nu=\lim_{r\to\infty} r^{-1}\ln|\mathrm{trace}\mathbf{Q}|$ as in Ref. 21.

[23] M. A. Wolfson and S. Tomsovic, "On the stability of long-range sound propagation through a structured ocean," J. Acoust. Soc. Am. **109**, 2693–2703 (2001).

Horizontal array beamforming in an azimuthally anisotropic internal wave field

Steven Finette[a] and Roger Oba
Acoustics Division, Naval Research Laboratory, Washington DC 20375

(Received 30 August 2002; revised 11 April 2003; accepted 11 April 2003)

A numerical study of beamforming on a horizontal array is performed in a shallow water waveguide where a summer thermocline is perturbed by a time evolving realization of an internal wave field. The components of the internal wave field consist of a horizontally (azimuthally) isotropic, spatially homogeneous contribution, and a horizontally anisotropic, spatially inhomogeneous component. These terms represent a diffuse (''background'') internal wave field and a localized solitary wave packet, respectively. Conventional beamforming is performed as a function of time while the internal wave field evolves throughout a computational volume containing the source–receiver paths. Source–receiver orientation with respect to the azimuthally anisotropic component has a significant effect on the beamformed output. When the source–receiver configuration is oriented approximately parallel to the solitary wave crests, beam wander, fading, beam splitting and coherence length degradation occurs in a time-dependent manner as the solitary wave packet passes through the environment. Both horizontal refraction of energy and a time-dependent modal source excitation distribution are responsible for these beamforming effects. In cases where source–receiver orientation is not approximately parallel to the wave crests, these effects are substantially reduced or eliminated, indicating that an azimuthally selective perturbation of the acoustic field can be attributed to the wave packet. Modal decomposition of the acoustic field and single mode starting fields are used to infer that, for the source–receiver orientation along the wave crests and troughs, acoustic propagation is predominantly adiabatic. A modal phase speed analysis explains several features associated with the beamformed power. [DOI: 10.1121/1.1582441]

PACS numbers: 43.30.Re, 43.30.Bp, 43.30.Ft [WLS]

I. INTRODUCTION

In this paper we present a model of horizontal array beamforming in a nonstationary shallow water waveguide, where the sound speed field is both azimuthally (horizontally) anisotropic and time dependent. Such an environment is commonly found in continental shelf areas, where tidal forcing on stratified water at the shelf break converts barotropic tidal energy to baroclinic tidal energy, often resulting in the formation of solitary wave packets (aka solibores) that propagate away from the shelf-break.[1] Recent modeling studies of acoustic propagation through solitary wave packets[2,3] show that horizontal refraction (HR) of the acoustic field has a significant effect on the space–time structure of the received signal for particular configurations of source, receiver, and packet wave number vector. The wave packet represents an azimuthally anisotropic oceanographic feature in the water column, and the horizontal refraction of energy is responsible for both focusing/defocusing of acoustic energy and the degradation of horizontal acoustic coherence when propagation is predominantly along the solibore wave crests and troughs.[4]

These studies, and others discussed in the following paragraph raise questions concerning sonar performance in the context of horizontal array beamforming in an azimuthally anisotropic, time-dependent sound speed field.[5] Interest in both the detection and localization of acoustic sources in dynamic ocean environments have served as the motivation behind the development of data adaptive and model-based beamforming algorithms (e.g., Refs. 6, 7). Both approaches attempt to optimize the processor output by taking advantage of information concerning the acoustic field and waveguide environment. For a continental shelf environment, we provide a clarification of some physical effects upon which such processors can be based. We consider conventional beamforming on both the full field and individual modes to illustrate, by simulation, how solitary wave packets can affect beamforming performance. Knowledge of the physical effects associated with time-dependent horizontal anisotropy is critical for the development of data adaptive and model based beamforming approaches designed to exploit the environment described in this paper, but their development lies outside the scope of this work.

A number of studies have discussed the horizontal refraction of acoustic energy in ocean waveguides, and a brief review is presented here to provide some perspective on the results obtained in this paper. Both ray and wave-based methods have been applied to estimate azimuthal angle deviations associated with HR induced by oceanographic, rather than bathymetric, variability in both deep and shallow water. The results obtained differ by an order of magnitude, due to the specifics of the oceanographic feature studied, the assumptions/simplifications used in the calculations, and the particular source/receiver configuration relative to the feature

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

of interest. Significant differences in HR can occur between azimuthally isotropic and anisotropic oceanographic features. In addition, the oceanographic field is sometimes closely coupled to the local bathymetry, making it difficult to separate water column and bottom contributions to azimuthal refraction. It should also be mentioned that HR effects associated with horizontal and vertical sound speed gradients as well as currents are not independent,[8] though they are usually treated separately. Ray-based calculations of horizontal refraction through deep water mesoscale eddies and fronts (e.g., Refs. 9, 10) yield values of 2°–4°. A parabolic equation approach[11] predicted a value of about 0.7° for eddy refraction, a result consistent with that based on adiabatic mode considerations applied to the same eddy model[12] using the theory of horizontal rays and vertical modes.[13] The role of currents was not included in the above calculations. Applying an eigenvalue perturbation approach to the method of Ref. 13, expressions were derived for HR induced by both volumetric sound speed inhomogeneities and bathymetric variations in the presence of currents.[14] More recently, HR due to internal gravity waves has been of interest. For an environment consisting of both a mesoscale eddy and an azimuthally isotropic (''background'') internal wave field described by a Garrett–Munk spectrum, an analysis using the horizontal ray/vertical mode approach concluded that the internal wave field made an insignificant contribution to HR, relative to the contribution due to the eddy.[15] On the other hand, for shallow water environments where azimuthally anisotropic solibores are often generated by tidal flow, the corresponding anisotropic correlation lengths allow for significant horizontal refraction of energy under certain conditions. These conditions are discussed next. An analysis of acoustic propagation along the wave crests, i.e., normal to the propagation vector of the solibore, indicates that HR can induce angular deviations of several degrees. This result has been discussed using horizontal ray/vertical mode calculations[2] and numerical simulations with two-dimensional (2-D)[2] and 3-D[3–5] parabolic equations. Other estimates of HR, based on an oblique transmission path through solitary wave packets during the Barents Sea Stationary Acoustic Range Experiment,[16,17] indicate angular deviations of about 0.1°–0.2°. A horizontal ray/vertical mode study of propagation through planar solitary wave packets predicts that when the acoustic propagation track forms an azimuthal angle greater than about 70° with the internal wave propagation vector, HR of acoustic energy will play a significant role,[18] a result confirmed by 3-D parabolic equation calculations used in conjunction with an energy gradient method.[3] The implication is that a three-dimensional model (i.e., a model that includes azimuthal coupling) will be needed to properly describe acoustic propagation under this condition. In related work, the azimuthal dependence of the acoustic field propagating through solitary wave packets has been studied using N×2-D parabolic equation simulations for single frequency[19,20] and pulse-synthesized acoustic fields,[21,22] as well as through a 2-D normal mode approach.[23] In these latter papers internal wave-acoustic wave resonance conditions, intensity and scintillation fluctuations, as well as time-of-arrival variations induced by the anisotropy of the internal wave packet are addressed; HR issues are not addressed because they neglect azimuthal coupling of the acoustic field. The azimuthal dependence of the resonance condition for a time-independent wave packet is correlated with strong angular dependence of energy transmission (e.g., Ref. 20). Pulse arrivals on a vertical array are shown to be strongly influenced by the orientation of the acoustic track with respect to the packet propagation direction; both the pulse shape and pulse travel time are affected.[21] In summary, a number of modeling studies predict that horizontally anisotropic wave packets have a significant effect on the azimuthal distribution of intensity and phase of an acoustic field. When acoustic propagation is directed primarily along the packet's wave crests and troughs, horizontal refraction assumes an important role.

The emphasis here is on propagation along the solibore wave crests and troughs, where HR is significant, though other propagation directions are also considered. The model environment and simulation technique are described in Sec. II. Plane wave beamforming on horizontal arrays using the full three-dimensional acoustic field is considered in Sec. III for a point acoustic source emitting at 400 Hz. The effect of source–receiver orientation on beam power is considered and a simple method of estimating horizontal coherence length using beam power is illustrated. Modal decomposition of the field and single mode starting fields are applied in Sec. IV to further clarify the relationship between the internal wave field, horizontal refraction, and beamforming effects. A summary and conclusions are presented in Sec. V.

II. MODEL DESCRIPTION

Both the environmental and acoustic propagation models used in the following sections have been described recently[3,24] and only a brief overview will be presented here. The four-dimensional $(\mathbf{r},t)=(x,y,z,t)$ space–time dependence of the sound speed field is represented by a functional of the isopycnal distributions corresponding to spatially diffuse $\eta_D(\mathbf{r},t)$ and spatially localized $\eta_S(x,z,t)$ internal wave fields,

$$c(\mathbf{r},t)=\overline{c}(x,z)+\delta c(\mathbf{r},t) =\overline{c}(x,z)+f(\eta_D(\mathbf{r},t)+\eta_S(x,z,t)). \quad (1)$$

In Eq. (1), $\overline{c}(x,z)$ describes the mean time-independent sound speed distribution containing range and depth variations but no cross-range dependence, while the time-dependent perturbation field is given by $\delta c(\mathbf{r},t)$. The perturbations are bounded as $[\delta c/\overline{c}]_{\max}\sim 10^{-2}$. The function $\eta_D(\mathbf{r},t)$ represents a three-dimensional realization of a spatially diffuse (''background'') internal wave field formed by a horizontally (azimuthally) isotropic, linear superposition of plane waves whose amplitudes are weighted by a shallow water displacement spectrum. The depth dependence is specified by the first internal wave mode obtained by solving an eigenvalue problem, given an estimate of the unperturbed density profile. The first mode typically dominates the vertical structure of the internal wave field in continental shelf environments. A dispersion relation linking spatial and temporal frequency evolves this field component in time. The

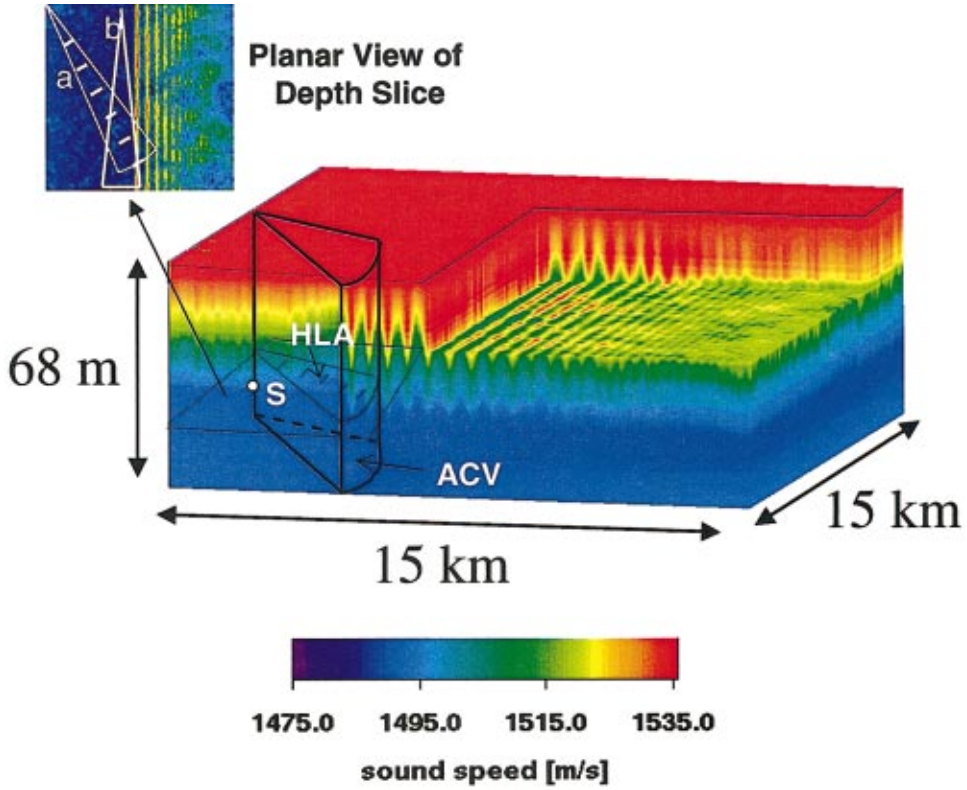


FIG. 1. An example of an environmental snapshot illustrating the water column sound speed distribution composed of a summer thermocline, spatially diffuse and spatially localized internal wave contributions. An acoustic computation volume (ACV) is placed in the environment with a source (*S*) at depth of 30 m. The horizontal slice through the ACV contains a sector with horizontal line arrays (HLA), which is also shown in plan view in the inset for two different azimuthal angles.

spatially localized contribution, $\eta_S(x,z,t)$, describes a solitary wave packet traveling in the x direction with a linear wave front (as viewed from the surface), the same depth dependence on the first mode, and an isopycnal recovery phase that returns the thermocline to its unperturbed depth after the packet passes through the region of interest. A linear wave front is a good approximation over the spatial scales ($\sim$10 km) considered in this paper. The functional form of the thermocline depressions comprised by the packet is given by the dnoidal elliptical function, for which the leading depressions have a secant-squared shape and the trailing depressions are approximately sinusoidal. The phase speed of the background field is approximately 0.5 m/s, while the speed of the leading edge of the packet is approximately 0.7 m/s. Wave packets are dispersive since the phase speed is a function of the depression amplitude. The maximum standard deviation of the background sound speed perturbations is 5.2 m/s and occurs at a depth of about 25 m; the leading solitary wave depression causes a maximum perturbation of 12–15 m/s. Both the time-independent sound speed $\overline{c}(x,z)$ and the internal wave field parameters are estimated from data obtained during the SWARM 95 experiment performed on the continental shelf off the New Jersey coast.[24]

An example of an environmental snapshot, acoustic computation volume (ACV) and a set of horizontal line arrays (HLA) are shown in Fig. 1. The snapshot represents the sound speed environment at a fixed time. Part of the volume is cut out so an interior section of the internal wave-induced sound speed variations can be seen. Full widths of the leading thermocline depressions measured at half-maximum are about 250–300 m with peak-to-peak distances of approximately 500 m. As mentioned above, the depth dependence is determined by the first internal wave mode. Since our purpose in this study is to focus on volume interactions, the bottom contribution is de-emphasized by choosing range-independent bathymetry and uniform bottom properties. A sound speed of 1700 m/s is chosen at the water/sediment interface (not illustrated); this speed increases linearly to 1750 m/s at 100 m depth (38 m into the sediment). Sediment density is chosen to be twice the water density, and an absorbing boundary is used at the bottom of the 38 m sediment layer.

A set of environmental snapshots represents the temporal evolution of a single realization of the sound speed distribution. This set is used as input to a three-dimensional parabolic equation that propagates the acoustic field from a 400 Hz point source (*S*) to the horizontal line arrays. Both the source and receiver arrays are fixed in space and are positioned within the wedge-shaped computational volume illustrated in Fig. 1. The source is positioned at array broadside for source/receiver ranges between 2.5 and 10 km. The computational volume can be rotated and translated to include different ''looks'' at the sound speed field. In Fig. 1, the array configuration is illustrated for a single depth slice within the computational volume and this slice is shown in the inset as a plan view. The inset depicts two array configurations, *a* and *b*, representing azimuthal angles of $\phi=\hat{\phi}=60°$ and $\hat{\phi}=90°$, respectively; these angles are measured between lines corresponding to array broadside and the propagation vector for the solitary wave packet. The beam power distribution on a horizontal array is a function of modal phase and group speeds for off-broadside sources; these speeds are generally not equal, even for a single propagating mode. As a consequence, the interpretation of the relationship between beam power maxima and source bearing is

simplified in the case where the source is located at array broadside,[25] and this is the source/receiver configuration chosen for this paper.

The three-dimensional parabolic equation code FOR3D[26] is applied here to compute complex acoustic fields. It is not only wide angle in elevation, but it also incorporates cross-range coupling via a narrow-angle approximation in azimuth. In cylindrical coordinates, the Helmholtz equation for the pressure field $p(r,\phi,z,\omega)$ is

$$\left[\rho\frac{\partial}{\partial r}\left(\frac{1}{\rho}\frac{\partial}{\partial r}\right)+\frac{1}{r}\frac{\partial}{\partial r}+\frac{\rho}{r^2}\frac{\partial}{\partial\phi}\frac{1}{\rho}\frac{\partial}{\partial\phi}+\rho\frac{\partial}{\partial z}\frac{1}{\rho}\frac{\partial}{\partial z}\right]p + k^2 p=0. \tag{2}$$

One may rewrite the pressure field in terms of the zeroth-order Hankel function of the first kind, $p=H_0^{(1)}(k_0 r)u$, where k_0 is the wave number associated with reference sound speed c_0. Equation (2) is approximated by a parabolic form in FOR3D using a pseudodifferential equation for acoustic field component $u(r,\phi,z,\omega)$:

$$\frac{\partial u}{\partial r}=(-ik_0+i\sqrt{k_0^2+Z+Y})u, \tag{3}$$

where $Z=k_0^2(n^2-1)+\rho\partial/\partial z(\rho^{-1}\partial/\partial z)$ with density $\rho=\rho(z)$, complex index of refraction $n=c_0/c(r,\phi,z)$ and $Y=(rk_0)^{-2}\partial^2/\partial^2\phi$. Equation (3) is integrated using an appropriate rational approximation for the square root of the differential operator. Acoustic propagation is simulated on a polar grid with the complex pressure field interpolated to a rectangular grid for beamforming computations. The three-dimensional acoustic field is computed with an angular (azimuthal) opening of 23°, of which only the central 13° is used. The 5° boundaries on either side were empirically determined to be sufficient to include the relevant horizontal scattering and exclude sidewall contamination of the central sector. This propagation wedge accommodated a set of 600 m horizontal arrays that are oriented perpendicular to a line connecting the source position and array center point and located between 2.5–10 km in range. Neumann sidewall conditions were applied on the sector boundaries with a two-dimensional solution. At 400 Hz, the gridding was 1.25 m in range, 1/16° in azimuth, and 0.2 m in depth; the same gridding is applied in Refs. 3–5.

We now manipulate FOR3D propagated fields to compute mode coupling. The function u can be decomposed into a normal mode series,

$$u=\sum_i \Psi_j(z;r,\phi)a_j(r,\phi), \tag{4}$$

where the jth vertical normal mode, Ψ_j, satisfies the eigenvalue problem $Z\Psi-\lambda\Psi=0$:

$$\left[\rho\frac{\partial}{\partial z}\frac{1}{\rho}\frac{\partial}{\partial z}+k_0^2(n^2-1)\right]\Psi_j-\lambda_j\Psi_j=0. \tag{5}$$

The eigenfunction Ψ is explicitly dependent upon z and implicitly dependent upon range r and azimuth ϕ through the density and index of refraction. The jth eigenvalue, $\lambda_j=\lambda_j(z;r,\phi)$, has similar implicit dependence. The mode coefficients $a_j(r,\phi)$ satisfy the jth modal decomposition relation,

$$a_j\equiv(u,\Psi_j)=\int_0^H u\Psi_j\rho^{-1}c_0^{-2}\,dz, \tag{6}$$

where H is the computational bottom and $(,)$ denotes an inner product. The algorithm in FOR3D typically uses a point starting field for the acoustic source. A slight modification of the model allows the use of the single lth mode starter field at $r=0$ defined by the boundary condition

$$u^l(r=0,z)=\Psi_l(z;r=0), \tag{7}$$

with the superscript denoting the starter mode. This starter field can be propagated through the environment and may then be filtered by the local jth mode at some range and angle: $(u^l(r,\phi),\Psi_j)$. Since FOR3D does not explicitly perform mode calculations, one may infer the degree of total mode coupling from the single mode starter. Define the dB scale modal coupling term, $C_j^l(r,\phi,t)$, by

$$C_j^l\equiv 20\log_{10}|(u^l,\Psi_j)|+10\log_{10}r, \tag{8}$$

where the subscript on C_j^l indicates the "received" mode and cylindrical spreading loss has been removed. This term represents "nonlocal" coupling between a mode l starter field with the acoustic field at some (r,ϕ,t); it therefore differs from the "local" mode coupling coefficient that relates coupling between adjacent range steps. In range- and angle-independent environments, the single mode starter field would propagate adiabatically as $\Psi_j e^{ik_j r}$; in this case, the negative of the self-coupling coefficients, $-C_l^l$, represent modal transmission loss. For the particular gridding selected, mode leakage in FOR3D for this range-independent environment is limited to at least -30 dB. That is, $C_j^l<-30$ if $l\neq j$ (<-40 dB for the first 5 modes at 400 Hz in this environment), while maintaining computational efficiency. Note that the parabolic equation uses a complex sound speed to model attenuation in the sediment and thus the modal coefficients will display some attenuation. Because of the choice of a relatively hard bottom and observed adiabaticity in the case examined, modal attenuation is not a major factor in the physical interpretation of the results in the following sections.

III. FULL FIELD BEAMFORMING

In this section we present some results of plane wave beamforming on large aperture horizontal arrays using the full acoustic field. The interpretation of the results is facilitated by our previous analysis of anisotropic propagation in this environment[3–5] and related work.[2] The arrays are processed independently; multiline or volumetric array processing is beyond the scope of this paper. The sound speed distribution is a function of both diffuse background internal waves and a spatially localized wave packet. The former is usually treated statistically, while the latter is typically considered as a deterministic wave. A simulated sound speed field should then be treated in a statistical manner though, for computational reasons, we consider here only a single representative realization of the sound speed field. This approach

serves as a practical trade-off between computational load considerations (3-D wave simulations) and accuracy in computing beamformed power, and allows us to focus on general trends associated with the relationship between the anisotropic internal wave field and array orientation. The environment is updated in one minute time increments and includes 200 consecutive snapshots of the 3-D environment. Since our purpose here is to consider only the effects of internal wave-induced volume variability on horizontal array beamforming, we also assume a noise-free environment. Isotropic ambient noise will not qualitatively affect the results, though anisotropic noise fields, e.g., from local discrete surface ships, could have a significant effect on the sidelobe structure. An oceanographic model including such noise fields is not considered here but would be important for system performance analysis. More realistic oceanographic models for describing the sound speed environment are under development.[27]

Given the complex acoustic field $p(r,\phi,z,\omega,t_q)$ for environmental snapshot t_q, the beamformed power $B_\Gamma(\theta)$ on a horizontal array is a function of both the bearing angle, θ, measured relative to endfire, and a set of parameters $\Gamma=\{r_i,\hat{\phi},z_j,\omega,t_q\}$. These parameters describe the range and depth of the (i,j)th array relative to the acoustic source, as well as its orientation $\hat{\phi}$, acoustic frequency, and snapshot number. Recall that the angle $\hat{\phi}$ denotes the central angle of propagation in the computational wedge, along the source–receiver path associated with array broadside. The beam power is given by the magnitude squared inner product between a vector whose elements are the complex acoustic fields measured on the array elements, $\overline{P_\Gamma}$, and a plane wave steering vector $\overline{S}(\theta_m)$:

$$B_\Gamma(\theta_m)=|\overline{S^\dagger}(\theta_m)\overline{P_\Gamma}|^2. \tag{9}$$

The variable θ_m is a discrete beam angle measured relative to endfire and † represents Hermitian conjugation. A wave front curvature correction[28] is applied to the steering vector since the computations are performed in the Fresnel region for this frequency and array length. Arrays are of equal length (600 m aperture, ~160λ at 400 Hz) with phone spacing of 1.25 m for a total of 481 phones. They are distributed in range with 60 m increments and in depth with increments of 5 m, starting 5 m below the ocean surface. The total number of beams is chosen to be twice the number of phones, and $\Delta\theta=(\theta_{m+1}-\theta_m)\approx 0.2°$.

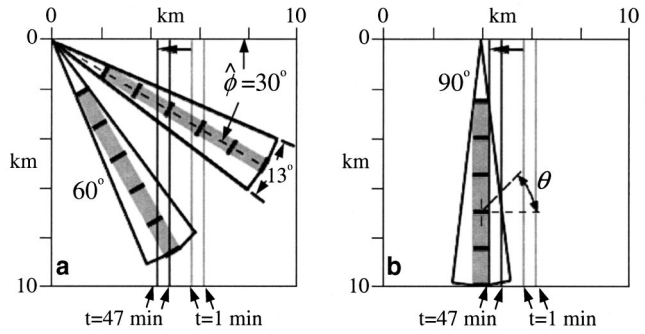


FIG. 2. Plan view of the acoustic propagation domains. (a) Acoustic propagation regions of 13° width within a 10 km square. The bearing of the centerline of the propagation path is relative to the top horizontal edge of the figure, as the 30° path illustrates. This edge is parallel to the solitary wave propagation vector. The gray rectangular regions in the center of the sectors are the locations of 600 m horizontal arrays transverse to the centerline (black segments shown at 1.5 km intervals) and positioned at ranges of 2.5–10 km from the acoustic source. Note that the actual array separations are 60 m in range. The pale gray vertical lines are the first two internal wave depressions at $t=1$ min (about 500 m apart) and the darker vertical lines are those depressions advanced leftward to $t=47$ min. The leading edge of the packet would reach the left edge of the domain at $t=135$ min. (b) The $\hat{\phi}=90°$ orientation of the arrays, with broadside "parallel" to the internal wave crests is shown, along with the definition of the bearing angle. The source position for the $\hat{\phi}=85°$ and 80° path coincides with that of 90°.

A. Bearing-time analysis

The complex acoustic field was computed in several ACVs for sets of horizontal arrays corresponding to azimuthal orientations of $\hat{\phi}=90°,85°,80°,60°,30°$. For reference, Fig. 2 illustrates planar views of the propagation domain for several of these orientations. Time is measured from an arbitrary origin. The gray regions in the sectors correspond to the areas in which the arrays are located; short line segments represent arrays. Vertical lines indicate the location of the first two solitary wave depressions for $t=1$ min and $t=47$ min. The bearing angle, θ, is defined relative to endfire in Fig. 2(b). Four environmental snapshots were selected for a special study using the 90° propagation path of Fig. 2(b). The sound speed distribution for these cases is illustrated in Fig. 3 at a depth of 30 m, the point source depth. Figure 3(a) shows the sound speed field in the range direction of propagation (i.e., along array broadside) through the spatially diffuse internal wave field at time $t=38$ min. The sound speed profile shows maximum variations of about 4 m/s at this depth. Figures 3(b)–(d) illustrate

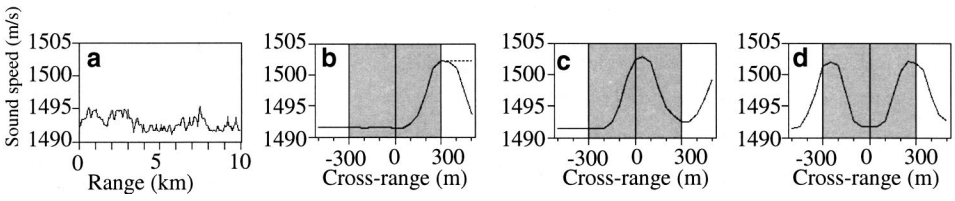


FIG. 3. Sound speed profiles along the radial and cross-range directions through the source position, at a source depth of 30 m for various times. (a) Sound speed at reference time $t=38$ min taken along a 10 km radial propagation path with only the random background internal wave field present. (b)–(d) Sound speed cuts in the transverse direction along a 600 m horizontal array: (b) for reference time $t=47$ min, just as the leading edge of the internal wave packet enters the region (indicated in gray) of the horizontal array. The solid curve shows the sound speed distribution of the packet; the upper dotted line is used for a comparison test case. (c) Time $t=54$ min, representing a "shadow" region where the soliton depression is centered on both the array center and on the acoustic source. (d) Time $t=61$ min, the "focusing" region, when the array center and acoustic source are located between first two internal wave depressions.

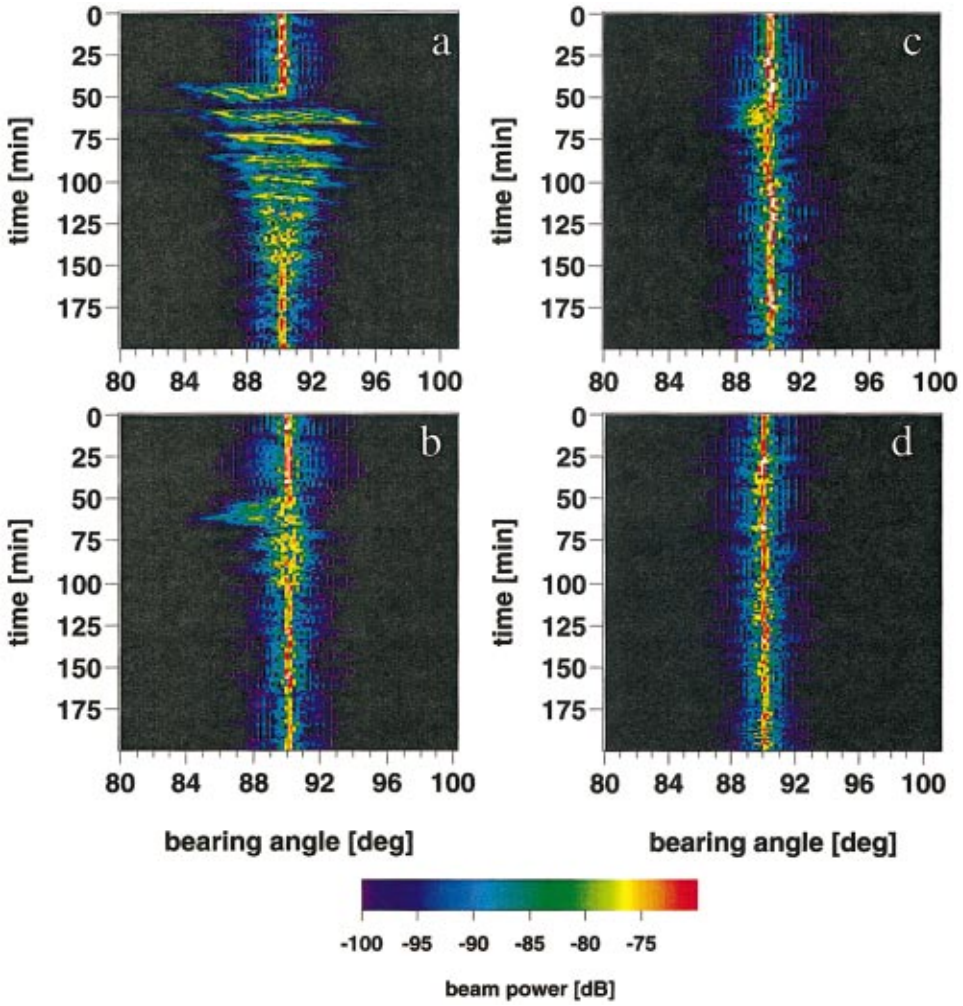


FIG. 4. Simulated bearing-time records for arrays located at ranges of 10 km and a depth of 35 m for different azimuthal angles between array broadside and the solitary wave propagation vector: (a) $\hat{\phi}=90°$, (b) $\hat{\phi}=85°$, (c) $\hat{\phi}=80°$, and (d) $\hat{\phi}=60°$. The bearing-time record for $\hat{\phi}=30°$ (not shown) is similar to that given in (c), (d).

the sound speed field in the transverse, i.e., cross-range direction, with the extent of the 600 m arrays shown in gray and 0 m indicating the array center point. The background isotropic variation is not readily visible in Figs. 3(b)–(d) because of the change in scale relative to Fig. 3(a). The solid curve in Fig. 3(b) represents the arrival of the first internal wave depression of the solitary wave packet ($t=47$ min). The dashed line to the right is the sound speed profile that will be used for a test comparison case in the next section. Note that the center of the arrays are still in the background region for this time. In Fig. 3(c), the center of the arrays nearly coincide with the center of the first internal wave depression ($t=54$ min) and Fig. 3(d) has the array bracketed by the first two internal wave depressions ($t=61$ min). The reader is referred to previous work,[3] where simulation showed that the environments corresponding to those in Figs. 3(b)–(d) were associated with significant distortion of the magnitude and phase of the acoustic field, resulting in horizontal refraction, coherence degradation, intensity shadowing, and focusing. Similar results were obtained using a different approach.[2]

Bearing-time records that span 200 min of temporal evolution of the sound speed environment are shown in Fig. 4, for an array located at a depth of 35 m and a range of 10 km from the acoustic source. These records are parametrized by azimuthal angles $\hat{\phi}=90°$, 85°, 80°, 60° in Figs. 4(a)–(d), respectively. Recall that the true source bearing is at broadside ($\theta=90°$) for all configurations of the source, array, and wave packet so that deviations of peak power from broadside can be interpreted as bearing errors caused by environmental variability. Beam wander is clearly a strong function of both azimuth, internal wave component, and time. When propagation is primarily along the wave crests [i.e., $\hat{\phi}=90°$ case in Fig. 2(b)], there is a sharp delineation between the time-dependent influences of the two internal wave components. This result is illustrated in Fig. 4(a) with reference to the time-dependent locations of the peak thermocline depressions shown in Fig. 3. For the first 43 min of the simulation, the spatially diffuse internal wave component is the only contributor to environmental variability. During this period, the size of the bearing errors is very small, around $\pm 0.2°$. As the solitary wave packet approaches the shaded region in Fig. 3(b) containing the array, horizontal refraction from the leading thermocline depression induces bearing errors. The maximum error is more than four degrees relative to broadside, with the initial deviation coming from the direction in which the packet is incident. From viewing a number of environmental snapshots and the corresponding acoustic fields, it can be inferred that the magnitude and direction of the bearing excursions relate to the quasiperiodic structure of the packet, depending on the location of the depressions relative to the array center, depression width, and peak to trough

sound speed difference. The spatially diffuse internal wave component, which is statistically isotropic, cannot account for these systematic bearing shifts, though it does contribute a small random variation to both the bearing variations and beam power variability. The amplitude of the thermocline depressions falls off significantly toward the back of the packet, where the thermocline is smoothly returning to its unperturbed value. After that time, occurring at about $t = 150$ min for this azimuth, the diffuse internal wave field again plays the dominant role in acoustic variability and both the bearing errors and beam power fluctuations are on the same scale found for times less than $t = 43$ min. Though not illustrated here, beam wander tends to increase with range from the source. Horizontal arrays positioned closer to the source experience less redirected energy from the wave packet, since any acoustic wave incident upon the thermocline trough will be above the critical angle and tend to propagate through the depression rather than be refracted back toward the array. Figure 4(a) is qualitatively representative of beam wander for arrays positioned at ranges greater than ~4 km from the source. In addition, beam power is modulated in range due to modal interference caused by the propagation of several vertical modes.

For a 5° shift in azimuth Fig. 4(b) shows that the beam wander is significantly suppressed. This damping is a strong function of azimuth, as illustrated in Figs. 4(a)–(d) and is consistent with theoretical and numerical predictions of refraction through a wave packet.[2,3,18] For times between approximately 50 and 70 min, there are elevated side lobes formed for the 85° and 80° azimuthal cases [Figs. 4(b) and (c)]; the beam wander is typically limited to less than 0.6°, though a 2° deviation can be seen in Fig. 4(b). The largest fluctuations of the peak power occur when the acoustic field propagates approximately along the wave crests/troughs, and they are present after each large bearing excursion in Fig. 4(a). These temporal variations in received power for this array/wave packet configuration are due to two factors. One is related to the creation of a shadow zone at the array caused by increased scattering from the sound speed perturbations as they pass over the source position, redirecting energy away from the array. The second factor is range-dependent focusing of acoustic energy between the depressions. The scattering is enhanced because of a redistribution of modal energy at the source, due to the change in the sound speed profile associated with the thermocline depression. Focusing is caused by horizontal refraction of acoustic energy between pairs of thermocline depressions when the source is located between a pair of depressions.[2,3,5] Smaller fluctuations in beam power are seen in Fig. 4(a) on the broadside beam for $t < 43$ min and $t > 150$ min, and are caused by sound speed variations induced by the background internal wave field.

To further clarify the influence of internal waves on beamforming, plots of beam power as a function of bearing angle are shown in Fig. 5 for several time cuts through Fig. 4(a). The figure illustrates beam power for the environmental snapshots shown in Fig. 3, corresponding to $t = 38$, 47, 54, and 61 min. These snapshots were chosen to illustrate acoustic field variability caused by qualitatively different local sound speed environments and are representative of results

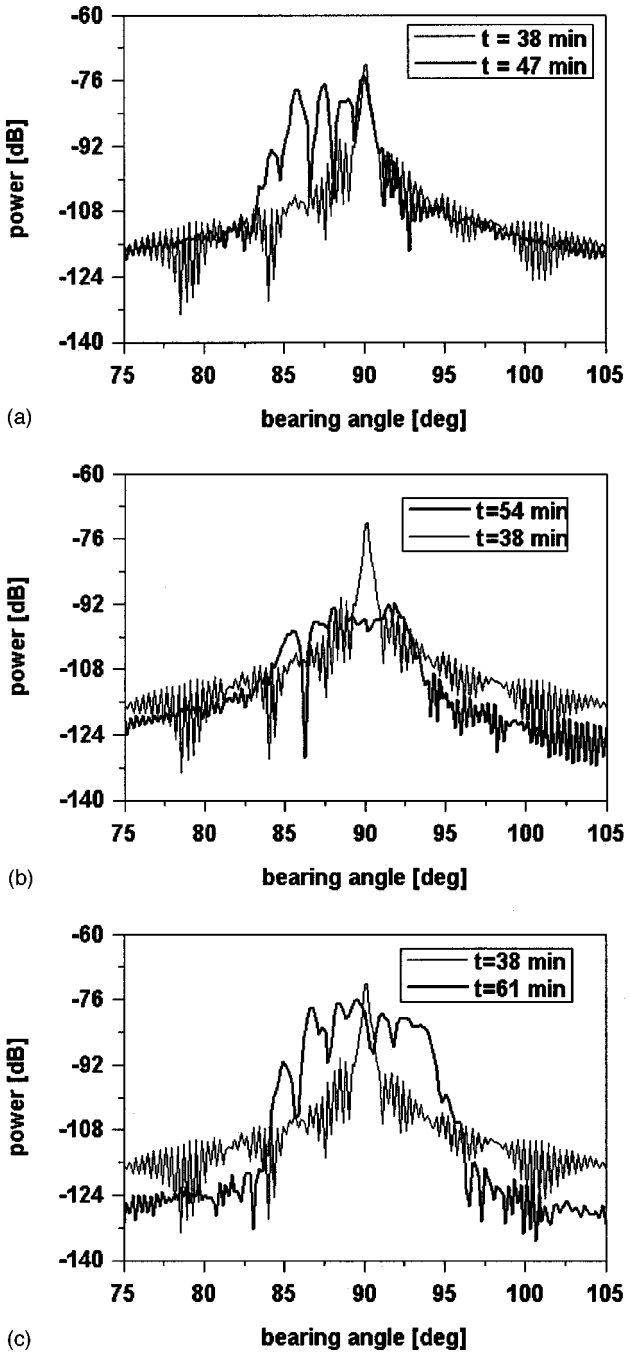


FIG. 5. Beam power for selected environmental snapshots: (a) $t = 47$ min, (b) $t = 54$ min, and (c) $t = 61$ min. The beam power at $t = 38$ min is given in each plot as a reference, corresponding to a typical beam power distribution when the internal wave packet was not influencing the propagation.

for other depths, though there are depth-dependent variations associated with the modal structure of the acoustic field. Only the background internal wave field is present at $t = 38$ min and the peak power correctly occurs at broadside for that time. This curve is reproduced on all plots in Fig. 5 for comparison purposes. Note that the background internal wave field induces very little beam shift (± 0.2–0.4°) from broadside over the first 43 snapshots. During this period, the solitary wave packet does not influence the acoustic field in the computational volume and the beamwidth at half-maximum is consistent with an aperture-limited array in this

environment. By $t=47$ min, the first thermocline depression has entered the computational region and the peak depression is located some 300 m from a line connecting the acoustic source and the array center [see Fig. 3(b)]. The peak power at broadside has dropped by about 4 dB [Fig. 5(a)], and additional sidelobes have emerged, indicating arrivals from different bearings in the direction of the leading thermocline depression. The strong sidelobes are due to the horizontal refraction of energy from the leading depression. Several of these off-broadside arrivals have levels comparable to the broadside peak with the result that, due to refraction, the total beam power received by the array has increased over that received at $t=38$ min. The physical significance of these strong sidelobes will be discussed in the next section, where the results from modal analysis are considered. For Fig. 5(b), the first thermocline depression is approximately centered over the source position and array center at $t=54$ min. In this case, beam power has dropped 24 dB at broadside and some of it is redistributed over bearing, within $\sim\pm5°$ of broadside. The rest of the scattered energy falls outside the array aperture. The uncertainty in bearing is due to significant scattering of the acoustic field from the higher sound speed in the thermocline depression to the lower sound speed region between the two depressions. In Fig. 5(c), corresponding to the environment illustrated in Fig. 3(c) at $t=61$ min, the packet has moved so that the acoustic source and array center are now bracketed by the first two thermocline depressions. The beam power in this case is significantly enhanced within $\sim\pm5°$ of broadside relative to the levels obtained at time $t=54$ min [Fig. 5(b)], and the peak levels are just a few dB below the background comparison case. Source bearing is affected by beamsplitting between three to four possible source-bearing estimates at this time. Additional acoustic energy, previously directed away from the array is now redirected toward the array by refraction from the pair of thermocline depressions, creating a natural oceanographic (horizontal) ducting or focusing effect at the frequency and range combination studied. While the environment appears rather symmetric with respect to the sound speed profile at the source location in Fig. 3(d), the acoustic field is also propagating through a background internal wave field that lacks this local symmetry; hence, the beam power is not necessarily symmetric with respect to array broadside in this case.

B. Estimation of coherence length for horizontal arrays

A number of approaches have been proposed to estimate the influence of ocean waveguide environments on transverse (horizontal) coherence in shallow water. An early experimental attempt using the aperture correlation coefficient considered winter conditions and emphasized decorrelation due to sea surface variations.[29] A model using horizontal rays and vertical modes indicated that the coherence of higher modes was more sensitive to the bottom and surface variability, while the volume variability represented the major cause of coherence loss for the first few modes.[30] The effect of sea surface scattering on spatial coherence in shallow water was emphasized in a modal formulation of the transport equation.[31] High-frequency (4–150 kHz) phase variability on horizontal arrays was considered in the context of imaging in Ref. 32, where it was found that (isotropic) internal waves and turbulence have quite different effects on spatial coherence. These approaches for estimating coherence, and others not discussed here, do not consider the azimuthally anisotropic, nonstationary oceanography emphasized in this paper.

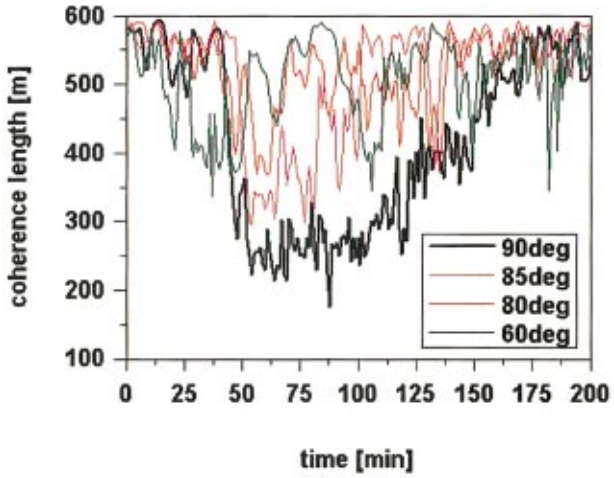


FIG. 6. The time dependence of coherence length for arrays located at a depth of 35 m and a range of 10 km from a 400 Hz source. The curves specify coherence length as a function of azimuth, with the acoustic source located at broadside for each azimuth.

Since beam wander and beam power variability have a significant dependence on both time and the horizontal anisotropy of the internal wave field, it is expected that coherence length should also be time dependent and sensitive to array orientation with respect to the direction of propagation of the wave packet. A simple estimate of coherence length as a function of time and azimuth is obtained as follows, using an approach similar in spirit to that in Ref. 33, where it is noted that the additional degrees of freedom present in the measurement of array signal gain offer an advantage over methods using the magnitude squared coherence function directly in estimating horizontal coherence. For a given snapshot, define the average element acoustic signal level over an array to be

$$\mathrm{SL}_{\mathrm{ph}}=10\log_{10}\frac{|\overline{P_\Gamma}|^2}{N}, \tag{10}$$

where N is the total number of hydrophones. The signal gain of the array, SG, is given by $\mathrm{SG}=\mathrm{SL}_a-\mathrm{SL}_{\mathrm{ph}}$, where the signal level for the array, SL_a, is defined by the relation $\mathrm{SL}_a=10\log_{10}(B_{\mathrm{peak}})$, with B_{peak} representing the peak beam power. The signal gain under ideal plane wave conditions, $\mathrm{SG}_{\mathrm{ideal}}$, is just $\mathrm{SG}_{\mathrm{ideal}}=20\log_{10}N$. Using the relations for SG and $\mathrm{SG}_{\mathrm{ideal}}$, one can relate the effective number of hydrophones, N_{eff}, to the signal gain. An estimate of the effective coherence length, L_{eff}, can be expressed in the form $L_{\mathrm{eff}}=(\lambda/2)10^{(\mathrm{SG})/20}$, where λ is the acoustic wavelength and a phone separation of $\lambda/2$ is assumed. Examples of L_{eff} are presented in Fig. 6 for arrays at the 10 km range, depth 30 m, and several azimuthal orientations. There is significant tem-

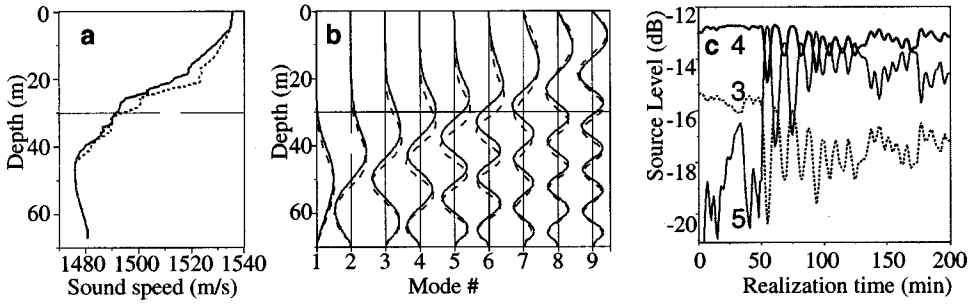


FIG. 7. Sound speed profiles, corresponding modal shapes, and time-dependent modal excitation levels at the source. (a) The sound speed profiles at the source location for the $\hat{\phi}=90°$ acoustic path. The solid curve corresponds to time reference $t=38$ min with the relatively unperturbed thermocline between 10 and 25 m. The dashed curve refers to the advected thermocline at $t=54$ min. (b) Solid vertical curves are the normal mode shapes at $t=38$ min. A horizontal line at 30 m depth references the point source depth and indicates the relative modal excitation, dominated by mode 4. These modes are very similar to source modes found at times 47 and 61 min [cf. Figs. 3(b) and (d) at 0 m cross-range]. The dashed curves are mode shapes for reference time $t=54$ min, corresponding to the time that the peak of the first wave packet depression crosses the source [cf. Fig. 3(c) at 0 m cross-range]. Modes are normalized separately for each time reference. (c) Modal source levels as a function of time for mode 4 (dark solid), mode 5 (light solid), mode 3 (dotted).

poral variability in coherence length throughout the 200 min of temporal evolution of the environment for each orientation of the array with respect to the direction of propagation of the wave packet. The initial degradation in coherence length is linked to the wave packet location with respect to the source/array propagation path. When the first thermocline depression begins to refract acoustic energy onto the array, the coherence length drops sharply, and this occurs at different times depending on azimuth (see Fig. 2). The largest drop in horizontal coherence length occurs for propagation along the wave crests. The level also drops to a minimum value in this case. A recovery phase then occurs, more gradually for the 90° orientation, as the depression amplitude decreases toward the back of the wave packet and the thermocline returns to its unperturbed position. For the 60° orientation, the packet influences the acoustic field at the array at an earlier time, around $t=15$ min, causing a drop in coherence length that then oscillates in time before settling down to its initial level. The earlier drop in coherence for this orientation is due to the simulation geometry. An analysis of coherence length measurements for several shallow water environments is presented in Ref. 33. The environments considered in Ref. 33 are substantially different from ours so that a comparison is not attempted. Our purpose here is to describe only the internal wave effects on coherence length; one would suspect lower values if, for the same water column environment, our estimates included other sources of coherence degradation such as scattering from rough waveguide boundaries. The point to be made here is that the wave packet significantly degrades coherence length for array broadside oriented along the wave crests, and that the degradation is quite sensitive to the horizontal anisotropy of the internal wave field.

IV. MODAL ANALYSIS

In this section we discuss results of modal calculations for acoustic propagation along the 90° azimuth, i.e., for propagation along the wave crests of a solitary wave packet. This orientation was selected because the most pronounced beamforming effects on the full field were found for this case. Although beamforming calculations are performed with a point source at 30 m depth, it will be demonstrated through examples that observations discussed in the previous section can be explained by the behavior of individual modes. Combinations of the modes can then be used to analyze the point source data. Single mode results show sharp features from horizontal refraction, which are less obvious in modal combinations.

The modal decomposition uses the normal mode code KRAKEN,[34] which, when appropriately gridded, has the same vertical differential operator as that used by FOR3D. A single mode starter field followed by modal decomposition of the propagated field at various receiver ranges is used to check the suitability of these modes.

The solid curve in Fig. 7(a) describes the sound speed distribution at the source position at $t=38$ min, when only the background internal wave field is present to perturb the thermocline. The thermocline is located between 10 and 25 m depth and the presence of the solitary wave packet advects it downward, as indicated by the dashed curve, computed at $t=54$ min. For a source depth of 30 m, the modal distribution at the source location for these two snapshots is shown in Fig. 7(b). A significant shift in peak amplitudes to deeper water depths occurs for modes 1–6, as the thermocline undergoes advection. This change is indicated by the shift from the solid mode curves ($t=38$ min) to the dashed mode curves ($t=54$ min). For this source depth, mode 4 has a peak at 30 m for $t=38$ min, and the peak amplitude is reduced by the internal wave packet so that it is no longer the dominant mode at $t=54$ min. At $t=54$ min, mode 5 becomes dominant as the internal wave passes through the region. From the excitation levels determined from Fig. 7(b), one can infer the relative influences of various modes on the full field at various depths, neglecting phase information.

The time dependence of modal source levels is illustrated in Fig. 7(c) for dominant modes 3, 4, and 5. A redistribution of source level among the modes is seen as the internal wave field evolves in time. Mode 4 has the strongest contribution during the period before the solitary wave packet begins to influence the acoustic field. When the wave packet enters the region containing the acoustic source, mode 5 becomes dominant for a brief period near $t=54$ min. The source level for mode 5 then tends to oscillate out of phase with that for mode 4, while the source level for mode 3 tends

to oscillate in phase with that of mode 4 as the depressions pass by the source position. All the other modes have source levels that are 19 dB or lower, except for mode 7, which is −17.2 dB or lower.

Because of differences in modal phase speeds, these modes interact slightly differently with the environment, so that the strongest direction of refraction, for example, will vary per mode and the full field will be refracted in a manner that depends on the relative modal contributions. In the beamforming case discussed above, differences in spatial phase speed gradients induce mode-order-dependent horizontal refraction. Examples of spatial variations in modal phase speed that cause mode-order-dependent variations in horizontal refraction at time $t=47$ min are plotted in Fig. 8 for modes 3, 4, and 7. This allows us to decouple the *vertical* mode shape effects from refraction induced by variations in *horizontal* phase speed. The phase speed curve for mode 4 (middle curve in Fig. 8) has about twice as much variation as that of mode 3 at the bottom of the figure, implying that mode 3 undergoes significantly less refraction than mode 4, while modes 1 and 2 exhibit very little refraction and are not represented in the figure. The phase speed of mode 7 (upper curve) has twice again as much variation as that for mode 4. It rises by 9 m/s at the leading edge of the internal wave packet, similar to the change in the sound speed shown in Fig. 3(b). It was found that the highest variation in phase speed occurred for modes 4–9, corresponding to variations of no more than 1% of the typical phase speed (approximately 1500 m/s).

When propagation is along the wave crests, the time-dependent enhancement of beamformed side lobe structure illustrated in the full field case can be explained in terms of time-varying modal contributions to the full field. It will be shown later from an analysis of modal coupling that acoustic propagation along the wave crests is primarily adiabatic. Each mode then travels with its own phase speed throughout the propagation path and undergoes different amounts of horizontal refraction as it propagates to the array, depending on the local environment. When there is a significant difference in acoustic phase speed induced by the two internal wave components (cf. Fig. 8), varying degrees of mode-dependent refraction appear as increased power on different beams. This can be explained theoretically by noting that the beam power, $B_\Gamma(\theta_m)$, is proportional to the square of the Green's function for a point source. Using a modal representation of the Green's function in wave number space, one can show that the beam power peaks occur at maxima corresponding to poles of the Green's function. These poles occur for horizontal wave numbers corresponding to the modal wave number eigenvalues.[35,36] This result relates the bearing angles at which one receives a maximum response to the modal eigenvalues. Beyond the critical angle for each mode, the beam power should drop significantly. An estimate of the maximum critical angles for the modes can be made using ratios of the phase speeds in the background and maximum depression regions of the internal wave environment. As an example, we use phase speed estimates for the environment at $t=47$ min (Fig. 8), to compute the critical angles for modes 3, 4, and 7. These angles are plotted in Fig. 9 as

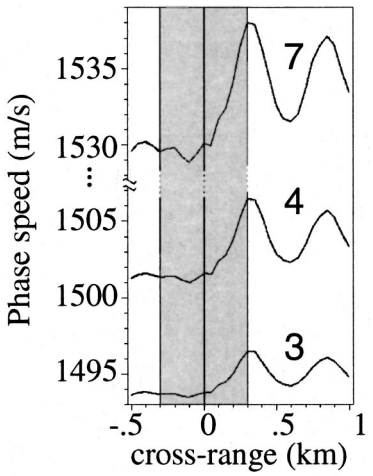


FIG. 8. Modal phase speeds as a function of cross-range at the source position for reference time $t=47$ min. The bottom, middle, and upper curves represent the phase speeds for modes 3, 4, and 7, respectively. The vertical gray band indicates the location of the 600 m array within the $\hat{\phi}=90°$ acoustic propagation path.

vertical dotted lines in conjunction with the modal power distribution for each of the three modes using this environmental snapshot. A good correlation between the estimates of the maximum critical angles and the position of the sharp drop-off in modal beam power is obtained.

Some representative examples of modal coupling coefficients, C_j^l, are shown in Fig. 10 for the environmental snapshots referenced in Fig. 3. A total of 16 modes were examined for mode coupling between mode starter fields at the source location and the acoustic field at some (r,ϕ,t) for these snapshots. Note the convention that the starter mode is underlined in this figure. The fourth mode starter field $l=4$ and the resulting redistribution of energy into local modes 3–5 are shown in Fig. 10(a) (from top to bottom) for $t=38$ min, a time when only the background internal wave component is present to perturb the sound speed distribution. Each plot in Fig. 10(a) shows the coupling between ranges 2.5–10 km and a cross-range from −300 to 300 m, as indi-

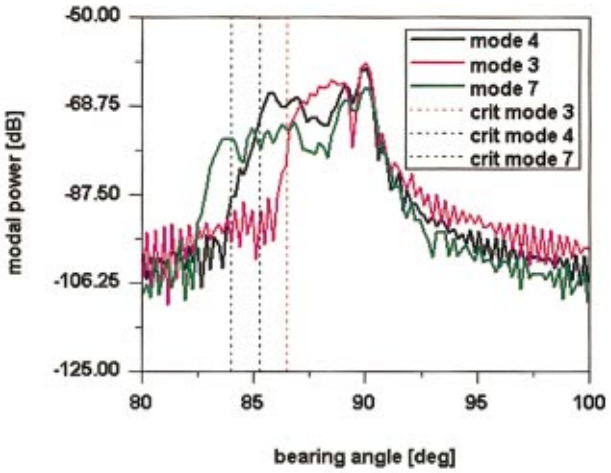


FIG. 9. Modal beamformed power for the environmental snapshot at $t=47$ min [cf. Fig. 3(b)]. The power distribution for modes 3, 4, and 7 are shown, along with critical angle estimates for each mode (dotted vertical lines) based on the modal phase speed plots given in Fig. 8.

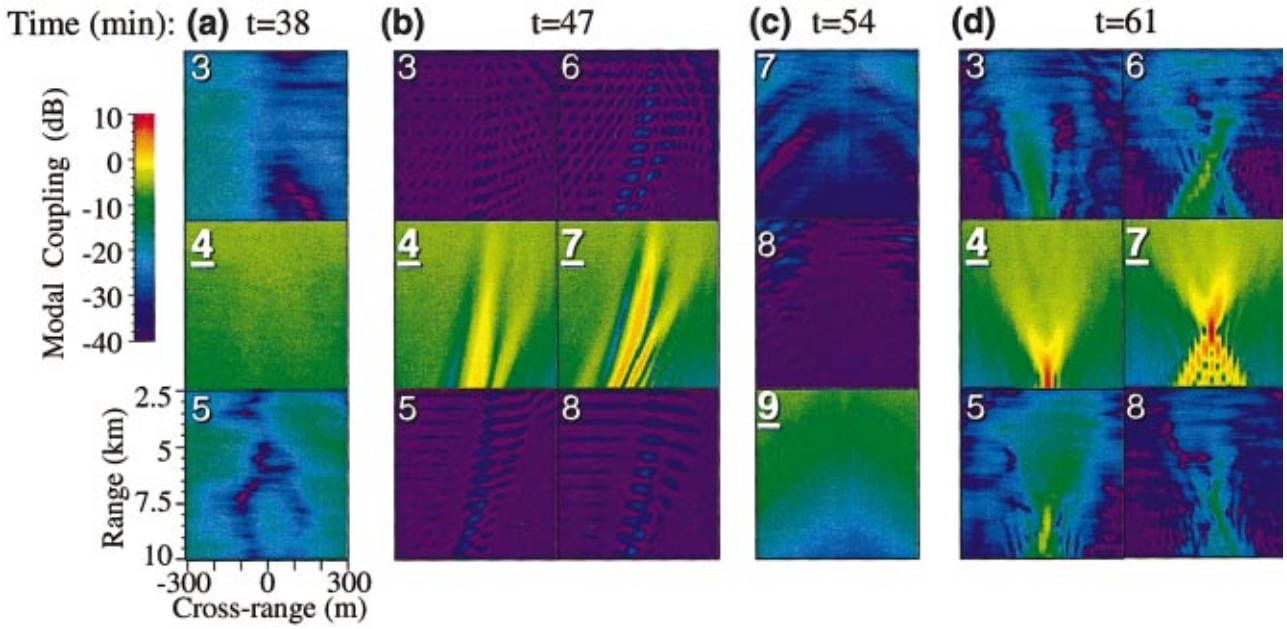


FIG. 10. Modal coupling amplitude plots. Single mode starter fields (mode order indicated by underlined number on vertical triple plots) are propagated to a range of 2.5–10 km over a total cross range of 600 m, as indicated in the bottom plot of (a). Each triple plot, at the indicated reference time, is filtered by the local mode shape (number indicated on an individual plot) for each range and cross-range. (a) At reference time $t=38$ min, with the background internal wave only, mode coupling to nearest neighbors is down by about 20 dB. Coupling to other modes (not shown) is down by 40 dB. (b) At reference time $t=47$ min, just as the internal wave packet enters the region including the horizontal array. These plots do not contain a background internal wave field. (c) At time $t=54$ min in the shadow region. (d) At time $t=61$ min, the focusing region, shows levels of coupling similar to those in (a).

cated in the bottom image. The same format is followed in Figs. 10(b)–(d). The self-coupling coefficient C_4^4 (center left) is nearly constant in azimuth. Plots directly above and below this figure represent ''off-diagonal'' coupling C_3^4, C_5^4, respectively, and this coupling is at least 10 dB lower than that for the self-coupling result. Additionally, the coupling of mode 4 to the other, undisplayed modes is at least 30 dB lower than the diagonal coupling term. Modes $l=1-16$ behave similarly for this reference time, and we infer that propagation is predominantly, though not exclusively adiabatic in this environment. The relatively low coupling is likely due to a rather weak transfer of energy between modes associated with the wave number resonance effect, in which coupling strength depends on the magnitude of the background wave number spectrum at wave numbers corresponding to pairwise differences in acoustic modal wave numbers. The resonance effect is described for 2-D background internal wave fields in Ref. 37 and has been extended to 3-D background fields.[38] Solitary wave packet resonance is considered in Ref. 39, with recent generalizations to horizontally anisotropic environments.[20,40]

Both mode 4 and mode 7 starter fields are considered in Fig. 10(b) for time $t=47$, just prior to the arrival of the internal wave packet at the source [see Fig. 3(b)]. In order to isolate the effect of the wavepacket on mode coupling, the background internal wave field has been removed by spatial averaging of c. [The case including the background internal wave field is very similar, but displays off-diagonal coupling at levels similar to Fig. 10(a).] In the absence of the random background field, the coupling to neighboring modes is 36 dB lower, both for $C_{3,5}^4$ and $C_{6,8}^7$. The results are quite similar for modes 1–16, indicating that acoustic propagation is primarily adiabatic along the wave crests and troughs. The energy propagated in the starter modes 4 and 7 are strongly refracted away from the front edge of the internal wave depression, which is located along the right-hand side of the image (not illustrated). As predicted by phase speed plots in Fig. 8, mode 7 exhibits stronger horizontal refraction than mode 4 and shows a bit more focusing of energy. Furthermore, if the sound speed is allowed to remain high on the backside of the leading solitary wave depression, as indicated by the straight dotted line on the right side of Fig. 3(b), the self-coupling is nearly identical to that shown in Fig. 10(b). Thus the horizontal refraction of modal energy seen in Fig. 10(b) is due only to the phase speed contrast on the left side of the leading depression, and the backside plays an insignificant role in the propagation. Modal coupling is not significant for propagation along the internal wave crests, and horizontal striations seen in off-diagonal coupling figures for $C_{3,5}^4$, $C_{6,8}^7$ occur at wavelengths appropriate to the difference of wave numbers for the modes, implying they are due to modal interference or leakage. A comparison to the case including the diffuse background internal wave field at $t=47$ min, not displayed here, shows a nearly identical pattern of horizontal refraction, but the peak received levels have transmission loss about 2 dB higher. The comparison also shows internal wave diagonal coupling at levels similar to that in Fig. 10(a). These results indicate the diffuse background perturbations are a comparatively weak scattering factor here, and this is consistent with the beamforming result discussed in the previous section. Horizontal refraction has a weak dependence on the random background contribution.

The remainder of the cases illustrated in Fig. 10 contain

both the diffuse background field and the solitary wave packet. Mode coupling results in Fig. 10(c) are determined from the sound speed distribution in Fig. 3(c), where the high sound speed thermocline depression is approximately centered on the array and source. This configuration causes a shadow region with much higher sound extinction and indications of refraction out of the central region. The shadow region is responsible for the overall lowering of beam power seen in Fig. 5(b). Also in this region, instead of nearest neighbor modal coupling, the dominant off-diagonal mode coupling for starter mode 9 is mode 7 rather than mode 8. Mode coupling coefficients in Fig. 10(d) corresponds to the sound speed field in Fig. 3(d), where the source is located in a region between the first two solitary wave depressions. Since the propagation for this environment is primarily adiabatic, the self-coupling coefficients can again be interpreted as modal transmission loss. Using starter modes 4 and 7, the focusing of energy is clearly seen at ranges of about 9 and 7.5 km, respectively. This range-dependent focusing is in accord with the higher horizontal refraction associated with mode 7. Indeed, the phase speed of mode 4 varies about 4 m/s, whereas the phase speed of mode 7 varies about 9 m/s. Thus, mode 7 also shows a finer interference pattern due to greater horizontal wave number angular variability. Mode coupling coefficients $C^4_{3,5}$ and $C^7_{6,8}$ are lower by about 10 dB, similar to the difference seen in Fig. 10(a). Note that the high coupling or ''focusing'' region of C^4_5 is slightly closer to the source, at about 8 km, than that for C^4_4 at about 9 km. This is consistent with energy being coupled from mode 4 to mode 5 by the background internal wave field for $r<5$ km; energy then propagates nearly adiabatically as mode 5, and it is focused according to the mode 5 phase speed differences. Conversely, C^4_3 shows a longer focusing range. These results are not as clear for the mode 7 starter field, because focal length as a function of mode order changes least near mode 7.

Focusing effects due to individual modal contributions quickly degrade as one deviates from the $\hat{\phi}=90°$ azimuthal path. Modal coupling remains at levels induced by the random background for the 60° path. Strong modal coupling is indicated for the 30° propagation paths. Additional calculations show that at various times the interaction with the internal wave packet can cause a mode 5 starter, for example, to be received at ranges 2.5–10 km, with the bulk of the energy in mode 3.

Horizontal focusing and refraction in the $\hat{\phi}=90°$ case can be analyzed on a mode by mode basis in the two horizontal coordinates (x,y) because the nearly adiabatic nature of the propagation, discussed above, permits an additional separation from the z coordinate. Consider the Helmholtz equation in three-dimensional Cartesian coordinates and assume that the z dependence satisfies an equation in the form of Eq. (5). The coefficients $a_j(r,\phi)\Rightarrow a_j(x,y)$ then satisfy the two-dimensional wave equation,

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\lambda_j\right]a_j=0, \tag{11}$$

where $+y$ is taken in the direction of propagation roughly parallel to r, and x denotes the cross-range direction. Neglecting the background internal wave field, the eigenvalue $\lambda_j=\lambda_j(x)$ results from a sound speed field $c(\mathbf{r},t)=\overline{c}(x,z)+f(\eta_S(x,z,t))$ that has only depth and cross-range variability for fixed t. The separation of variables gives $a_j(x,y)=A_j(x)e^{i\kappa y}$ and the cross-range eigenvalue equation,

$$\left[\frac{\partial^2}{\partial x^2}+(\lambda_j-\kappa^2)\right]A_j=0. \tag{12}$$

This expression is now an ordinary differential equation, and can be solved by one-dimensional methods. In particular, consider $t=61$ min, and vertical mode 7 within the gray-shaded focusing region of Fig. 3(d). In that region, the 7th vertical mode has the phase speed distribution $c_{\mathrm{ph},7}(x)$ over a cross-range of 600 m shown in Fig. 11. To create a canonical horizontal duct for comparison, phase speeds for $x>300$ m and $x<-300$ m have been replaced by $c_{\mathrm{ph},7}(x)=1537.28$ m/s (equal to the right-hand peak in the figure), resulting in the single ''well'' shown in Fig. 11. The actual focusing region is a leaky duct that would properly be modeled by a continuum of horizontal modes. The energy will gradually tunnel through the high sound speed regions in the cross-range direction as the range increases. The ''replacement'' phase speed gives a true duct with horizontally trapped modes, and can be solved numerically by Kraken using fluid half-spaces to the left and right of the well. The phase speed appears in Eq. (12) as $\lambda_7(x)=[2\pi f/c_{\mathrm{ph},7}]^2$. In this sound speed environment, Eq. (12) supports 17 horizontal modes $A_m(x)$ and eigenvalues κ_m^2, $m=1,2,...,17$. Associated phase speeds for the odd modes are indicated in Fig. 11 by horizontal lines. Since a mode is evanescent when $2\pi f/\kappa_m<c_{\mathrm{ph},7}(x)$, the length of the line segment in the well indicates the cross-range extent of a horizontal mode; for example, A_{17} only contributes significantly for $|x|<250$ m. The minimum phase speed $c_{7,\min}=\min[c_{\mathrm{ph},7}(x)]=1530.93$ m/s at $x=0$ has a maximum wave number $\kappa_{\max}=2\pi f/c_{7,\min}$. Hence, near $x=0$, the forward propagating so-

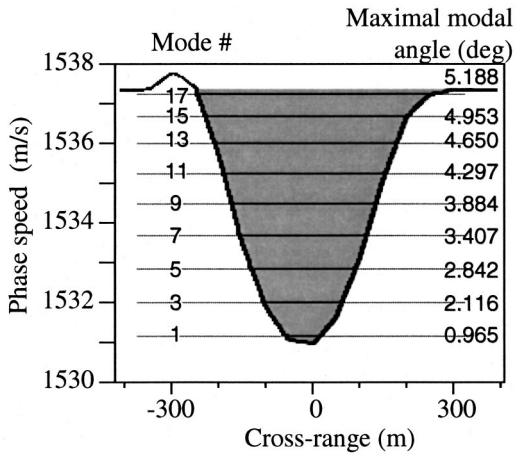


FIG. 11. Vertical mode 7 phase speed plot for 400 Hz, using the potential well approximation to the focusing region. The horizontal mode phase speed lines for the odd numbered modes indicate the horizontal extent of the modes within the well. The maximal modal angle limits the beam broadening for the trapped mode.

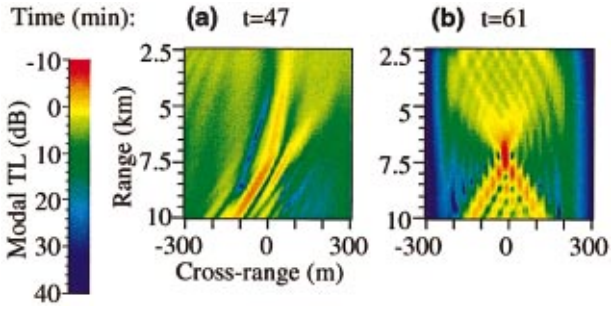


FIG. 12. Reconstruction of the modal transmission loss (negative of the self-coupling coefficient) using the vertical mode 7 phase speed to compute a cross-range modal solution. (a) Results for time $t=47$ min using the reflection coefficient method and Fourier transform. (b) Results for $t=61$ min using purely modal methods.

lution will behave like $A_m(x)e^{i\kappa_m y} \approx e^{i\kappa_m y}(\gamma e^{i\eta x} + \overline{\gamma} e^{-i\eta x})$, where γ is a complex constant and $\eta = \sqrt{\kappa_{\max}^2 - \kappa_m^2}$.

This solution for the azimuthal modes trapped in the duct behaves like the sum of a solution slightly deviated to the left by $\xi = a \tan(\eta/\kappa_m)$ and a rightward deviated solution of the same angle. Using the above analysis, we can make an estimate of the expected amount of beam broadening per mode. The solution will exhibit approximate beam broadening by the maximal modal angle ξ, which, in this example, varies from about 1° to 5.2°, as indicated in Fig. 11. This estimate corresponds reasonably well to the full field (multimodal) broadening at time $t=61$ shown in Fig. 5(c), with peaks appearing on either side of 90°, starting with variations of approximately ±1° and going up to about ±6°. A comparison with the actual mode 7 beamformed field at $t=61$ min (not illustrated) verifies this result.

Here we consider how the cross-range modes, obtained above for vertical mode 7, can be used to estimate the modal transmission loss $-C_7^7$. Figure 12 shows the transmission loss using two-dimensional reconstruction for the adiabatic horizontal modal coefficients $a_7(x,y)$ corresponding to vertical mode 7 at times (a) $t=47$ min and (b) $t=61$ min. The results can be compared to modal coupling shown in Figs. 10(b) and (d), as they correspond to the negative of modal transmission loss. The calculation in Fig. 12(a) uses a reflection coefficient method for the y solutions, similar to that found in Ref. 41. The resulting TL levels are again about 2–3 dB higher at the focusing region than in Fig. 10(b). The deflection is slightly higher and the transmission loss structure is more intricate when computed with the reflection coefficient method. In Fig. 12(b), the resulting transmission loss is the combination of the 17 horizontal mode solutions and these sum to a focus at about 7.3 km, a location slightly closer than the focusing range of 7.5 km in Fig. 10(d) that is computed with FOR3D. The peak intensity is about 2 dB higher in the modal calculation than the value found using FOR3D. Naturally, the horizontal mode interference pattern from FOR3D is not as clear as that found from the true modal calculation. The horizontal modal approximations, including the use of canonical ducts, permit refraction angle and focusing distance estimates that confirm the horizontal refraction interpretation of the propagation.

While the results discussed in this paper are for a source frequency of 400 Hz, several additional simulations were performed at other frequencies. The first mode exists down to approximately 25 Hz, and shows ducting within the slower sound speed regions down to that frequency for the $\hat{\phi}=90°$ azimuthal orientation. Horizontal focusing is supported as long as more than one horizontal mode exists, as one would expect from the modal interference interpretation given above. Therefore, use of the vertical modal phase speeds, in conjunction with the horizontal mode decomposition, provide a predictive tool at other frequencies in this orientation.

V. SUMMARY AND CONCLUSIONS

In this paper we describe how a time-dependent, azimuthally anisotropic sound speed field can affect horizontal array beamforming in a shallow water waveguide. The sound speed distribution is modeled in terms of two internal wave components. One contribution is azimuthally isotropic and corresponds to a background field, while the other is azimuthally anisotropic and represents an internal solitary wave packet. The isotropic internal wave field has relatively little effect on the beamformed power distribution; the beamwidth is consistent with an approximate diffraction-limited aperture when propagation occurs in an environment in which only this component is perturbing the mean sound speed field. The presence of a horizontally anisotropic wave packet alters this situation. When the wave packet is present, the distribution of beam power in bearing angle and time is strongly dependent on the source–receiver orientation with respect to the propagation vector of the solitary wave packet. Choosing the source–receiver orientation approximately parallel to the plane wave packet's crests and troughs causes significant time-dependent beam wander of up to four degrees off the true bearing (at broadside) for a range of 10 km. The beam wander decays in time over about two hours as the packet passes through the region containing the array. Beam broadening and the degradation of coherence length are maximized for this orientation, and horizontal refraction is responsible for these effects. They become strongly suppressed when the source/array configuration is rotated by 5°–10°. Modal decomposition of the acoustic field and single mode starting fields indicate that propagation is primarily, though not exclusively, adiabatic for propagation down the crests and troughs of the solitary wave packet. For adiabatic propagation, the depth-dependent modal contribution can be separated out of the acoustic field computation. In this case, a two-dimensional model for the cross-range modal contributions to the acoustic field for a given vertical mode allows us to estimate refracted energy and the location of the focusing region. Significant mode coupling from the internal wave packets only occurs at angles where the propagation path crosses the internal wave troughs and crests obliquely. The diffuse isotropic internal wave field allows a small amount of modal coupling at all propagation angles.

We have discussed horizontal array beamforming in a rather simple oceanographic setting, emphasizing only the effect of volumetric sound speed variability induced by internal gravity waves. Neither solitary wave front curvature,

bathymetric variations, ambient noise fields, or rough ocean boundaries are considered. In addition, the acoustic source and receiver array are fixed in space as the environment evolves. All of the above-mentioned items can affect the beamforming performance. While the results in this paper provide some clarification of the physics of anisotropic propagation that may impede beamforming operations, the significance of the results obtained here, in the context of sonar performance, will have to await a further analysis in which other potential sources of degradation are incorporated into the environmental model.

ACKNOWLEDGMENTS

This work was supported by funds from the Office of Naval Research. A grant of HPC computer time was obtained from the NRL. The authors wish to thank Dr. Stephen Wolf and Dr. T. C. Yang for several useful discussions, and the comments of the reviewers were most appreciated.

[1] J. R. Apel, L. A. Ostrovsky, and Y. A. Stepanyants, ''Internal solitons in the ocean,'' Johns Hopkins University Applied Physics Laboratory, Laurel, MD, Report MERCJRA0695, 1995.
[2] B. G. Katsnel'son and S. A. Pereselkov, ''Low-frequency horizontal acoustic refraction caused by internal wave solitons in a shallow sea,'' Acoust. Phys. **46**, 684–691 (2000).
[3] R. Oba and S. Finette, ''Acoustic propagation through anisotropic internal wave fields: Transmission loss, cross-range coherence, and horizontal refraction,'' J. Acoust. Soc. Am. **111**, 769–784 (2002).
[4] S. Finette and R. Oba, ''Horizontal coherence estimates for an internal wave dominated shallow water environment,'' in *Proceedings of the 5th European Conference on Underwater Acoustics*, edited by M. E. Zakharia, P. Chevret, and P. Dubail, Lyon, France, 2000, pp. 151–156.
[5] S. Finette and R. Oba, ''Acoustic field variability induced by internal solitary waves on a continental shelf,'' in *Acoustical Oceanography*, edited by T. G. Leighton, G. J. Heald, H. D. Griffiths, and G. Griffiths, 2001, Proceedings of the Institute of Acoustics, Vol. 23, pp. 321–326.
[6] J. Krolik, ''Matched-field minimum variance beamforming in a random ocean channel,'' J. Acoust. Soc. Am. **92**, 1408–1419 (1992).
[7] J. V. Candy and E. J. Sullivan, ''Model-based identification: an adaptive approach to ocean acoustic processing,'' IEEE J. Ocean. Eng. **21**, 273–289 (1996).
[8] R. F. Henrick, M. Jacobson, and W. Siegmann, ''General effects of currents and sound-speed variations on short-range acoustic transmission in cyclonic eddies,'' J. Acoust. Soc. Am. **67**, 121–134 (1980).
[9] W. H. Munk, ''Horizontal deflection of acoustic paths by mesoscale eddies,'' J. Phys. Oceanogr. **10**, 596–604 (1980).
[10] N. L. Weinberg and J. G. Clark, ''Horizontal acoustic refraction through ocean mesoscale eddies and fronts,'' J. Acoust. Soc. Am. **68**, 703–706 (1980).
[11] R. N. Baer, ''Propagation through a three-dimensional eddy including effects on an array,'' J. Acoust. Soc. Am. **69**, 70–75 (1981).
[12] M. V. Hall and M. A. Irving, ''Application of adiabatic mode theory to the calculation of horizontal refraction through a mesoscale ocean eddy,'' J. Acoust. Soc. Am. **86**, 1465–1477 (1989).
[13] H. Weinberg and R. R. Burridge, ''Horizontal ray theory for ocean acoustics,'' J. Acoust. Soc. Am. **55**, 63–79 (1974).
[14] Yu. A. Kravtsov, V. M. Kuz'kin, and V. G. Petnikov, ''Perturbation calculation of the horizontal refraction of sound waves in a shallow sea,'' Sov. Phys. Acoust. **30**, 45–47 (1984).
[15] A. G. Voronovich and E. C. Shang, ''A note on horizontal-refraction-modal tomography,'' J. Acoust. Soc. Am. **98**, 2708–2716 (1995).
[16] A. Yu. Shmelerv, A. A. Migulin, and V. G. Petnikov, ''Horizontal refraction of low-frequency acoustic waves in the Barents Sea stationary acoustic track experiment,'' J. Acoust. Soc. Am. **92**, 1003–1007 (1992).
[17] V. M. Kuz'kin, ''Horizontal refraction of acoustic waves at internal waves in a shallow sea,'' Acoust. Phys. **42**, 390–392 (1996).
[18] J. C. Preisig and T. F. Duda, ''Coupled acoustic mode propagation through continental-shelf internal solitary waves,'' IEEE J. Ocean. Eng. **22**, 256–269 (1997).
[19] B. G. Katsnel'son, S. A. Pereselkov, V. G. Petnikov, K. D. Sabinin, and A. N. Serebryanyi, ''Acoustic effects caused by high-intensity internal waves in a shelf zone,'' Acoust. Phys. **47**, 424–429 (2001).
[20] D. Rubenstein, ''Observations of cnoidal internal waves and their effect on acoustic propagation in shallow water,'' IEEE J. Ocean. Eng. **24**, 346–357 (1999).
[21] M. Badiey, Y. Mu, J. Lynch, J. Apel, and S. Wolf, ''Temporal and azimuthal dependence of sound propagation in shallow water with internal waves,'' IEEE J. Ocean. Eng. **27**, 117–129 (2002).
[22] K. V. Konyaev, D. E. Leikin, K. D. Sabinin, Yu. I. Tuzhilkin, and V. R. D'yachenko, ''Correlation between internal solitons and amplitude variations of a sound signal on a fixed path,'' Acoust. Phys. **44**, 476–485 (1998).
[23] V. G. Petnikov, B. G. Katsnel'son, S. A. Pereselkov, and K. D. Sabinin, ''Numerical modeling of sound field perturbations caused by internal waves in shallow water,'' *Proceedings of the International Conference on Shallow Water Acoustics*, Beijing, China, edited by R. Zhang and J. Zhou (China Ocean Press, Beijing, 1997), pp. 289–294.
[24] S. Finette, M. H. Orr, A. Turgut *et al.*, ''Acoustic field variability induced by time-evolving internal wave fields,'' J. Acoust. Soc. Am. **108**, 957–972 (2000).
[25] C. S. Clay, ''Array steering in a layered waveguide,'' J. Acoust. Soc. Am. **33**, 865–870 (1961).
[26] D. Lee and M. H. Schultz, *Numerical Ocean Acoustic Propagation in Three Dimensions* (World Scientific, Singapore, 1995).
[27] S. Finette, T. Evans, and C. Shen, ''Sub-mesoscale modeling of environmental variability in a shelf-slope region and the affect on acoustic fluctuations,'' in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer Academic, Dordrecht, 2002) pp. 401–408.
[28] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1993).
[29] P. Wille and R. Thiele, ''Transverse horizontal coherence of explosive signals in shallow water,'' J. Acoust. Soc. Am. **50**, 348–353 (1971).
[30] R. Zhu and D. Guan, ''Spatial horizontal coherence of sound in shallow water,'' J. Acoust. Soc. Am. **92**, 956–961 (1992).
[31] A. G. Sazontov, A. L. Matveyev, and N. K. Vdovicheva, ''Acoustic coherence in shallow water: theory and observation,'' IEEE J. Ocean. Eng. **27**, 653–663 (2002).
[32] F. S. Henyey, D. Rouseff, J. M. Grochocinski, S. A. Reynolds, K. L. Williams, and T. E. Ewart, ''Effects of internal waves and turbulence on a horizontal aperture sonar,'' IEEE J. Ocean. Eng. **22**, 270–280 (1997).
[33] W. M. Carey, ''The determination of signal coherence length based on signal coherence and gain measurements in deep and shallow water,'' J. Acoust. Soc. Am. **104**, 831–837 (1998).
[34] M. B. Porter, ''The KRAKEN normal mode program,'' Rep. SM-245, SACLANT Undersea Research Centre, La Spezia, Italy, 1991.
[35] W. A. Kuperman, M. F. Werby, K. E. Gilbert, and G. J. Tango, ''Beam forming on bottom-interacting tow-ship noise,'' IEEE J. Ocean. Eng. **10**, 290–298 (1985).
[36] G. V. Frisk, J. F. Lynch, and S. D. Rajan, ''Determination of compressional wave speed profiles using modal inverse techniques in a range-dependent environment in Nantucket Sound,'' J. Acoust. Soc. Am. **86**, 1928–1939 (1989).
[37] L. B. Dozier and F. D. Tappert, ''Statistics of normal mode amplitudes in a random ocean. I. Theory.,'' J. Acoust. Soc. Am. **63**, 353–365 (1978).
[38] C. Penland, ''Acoustic normal mode propagation through a three-dimensional internal wave field,'' J. Acoust. Soc. Am. **78**, 1356–1365 (1985).
[39] J. 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).
[40] B. G. Katsnel'son and S. A. Pereselkov, ''Resonance effects in sound scattering by internal wave packets in a shallow sea,'' Acoust. Phys. **44**, 684–689 (1998).
[41] L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer-Verlag, Berlin, 1991).

Robust time reversal focusing in the ocean

Seongil Kim,[a] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, and T. Akal[b]
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 14 March 2002; revised 5 March 2003; accepted 13 April 2003)

Recent time-reversal experiments with high-frequency transmissions (3.5 kHz) show that stable focusing is severely limited by the time-dependent ocean environments. The vertical focal structure displays dynamic variations associated with focal splitting and remerging resulting in large changes in focal intensity. Numerical simulations verify that the intensity variation is linked to the focal shift induced by phase changes in acoustic waves resulting from sound speed fluctuations due to internal waves. A relationship between focal range shift, frequency shift, or channel depth changes is illustrated using waveguide-invariant theory. Based on the analysis of experimental data and numerical simulations, methods for robust time-reversal focusing are developed to extend the period of stable focusing. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582450]

PACS numbers: 43.30.Vh, 43.30.Re, 43.60.Gk [RS]

I. INTRODUCTION

Time reversal[1–4] is a process of transmitting a received signal in a time-reversed order. It exploits the time-reversal invariance of the wave equation such that the retransmitted wave from a time-reversal mirror (TRM) converges back to the position of a probe source (PS) where the signal was generated. The fact that time-reversal processing (TRP) both spatially and temporally refocuses the acoustic energy in a complicated ocean environment suggests possible applications to future active sonar and underwater communication systems.

Since TRP utilizes the advantage of reciprocity in a static medium, the performance can be degraded in a nonstatic environment,[5,6] where the propagation conditions change during the time between the forward propagation of a probe signal and the backpropagation of the time-reversed signal. In recent experiments, the stability of the focal structure was investigated by repeatedly retransmitting a stored probe source pulse into the time-dependent ocean medium. While the focal structure was maintained over several hours to days with low-frequency (445 Hz) transmissions,[7] measurements with high-frequency (3.5 kHz) transmissions showed that the stable focus was limited to on the order of a few tens of minutes, implying that the effect of medium fluctuation is not trivial for such high-frequency transmissions.

The ability to maintain a stable focus is important in the applications of TRP. In underwater communications it is desirable to extend the period of reliable communications using a single probe pulse captured in order to increase the data rate by increasing the time intervals between the training probe pulses.[8] Similarly, in the acoustic barrier concept,[9] a disturbance between the TRM and the focus can be monitored continuously over an extended period of time without refreshing the TRM focus. The analogy between TRP and matched-field processing (MFP) also can provide useful guidance for developing more robust MFP methods.

Our objective in this paper is to develop an efficient method for robust time-reversal focusing in a fluctuating ocean environment. To achieve this, the experimental data are analyzed to extract the main features of focal stability. The detailed response of the time-reversal focus to sound-speed fluctuations driven by internal waves is then examined using numerical simulations. A simple theory based on waveguide invariants[10–13] can describe the relationship between focal stability and the changes in environmental conditions. Finally, based on the analysis of experimental and numerical simulations, methods for robust time-reversal focusing are developed to extend the period of stable focusing.

In the next section, focal resolution and stability of ocean acoustic time reversal are reviewed briefly. In Sec. III, time-reversal experiments and the measured focal stability in a fluctuating ocean are described. In Sec. IV, the effect of sound-speed fluctuations on the spatial focal structure is investigated using numerical simulations. In Sec. V, a theoretical analysis based on waveguide invariant theory is examined to describe the focal shift caused by sound-speed variations. In Sec. VI, robust time-reversal methods are explored using multiple constraint methods.

II. OVERVIEW OF OCEAN TIME-REVERSAL ACOUSTICS

Here, we briefly describe the concept of time-reversal processing, focal size, and the effect of sound-speed fluctuations on the focal structure. Additional details about the general concept of ocean TRP and the analogy with matched-field processing can be found in Refs. 4 and 14.

TRP involves two-way wave propagation between a probe source and a time-reversal mirror, as depicted in Fig. 1. The TRM refocuses wave energy back to the position where the signal was generated. The signal can be a man-made sound, natural noise, or scattered sound from rough ocean boundaries, etc. In the experiments, we used a PS to

[a]Presently S. Kim is with the Agency for Defense Development, Chinhae 645-600, South Korea. Electronic mail: sikim@add.re.kr
[b]This work was done while at the SACLANT Undersea Research Center, 19138 La Spezia, Italy. Current address: TUBITAK-MAN, Marmara Research Center, Earth and Marine Sciences Research Institute, P.K.21 Gebze, Kocaeli 41470, Turkey.

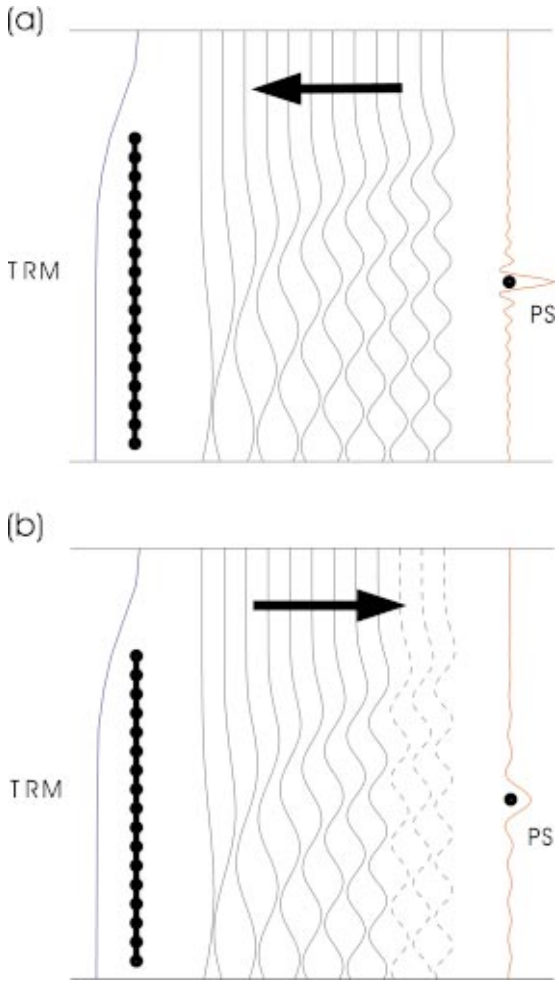


FIG. 1. Schematic of time-reversal processing: (a) a probe signal is excited and propagates to the TRM; (b) time-reversed signal propagates backward from the TRM.

generate an initial pulse. The excited acoustic field can be decomposed into modes and each mode propagates with its own group speed. In general, the lower-order modes propagate faster in shallow water and arrive earlier at the TRM. The received signal exhibits dispersion patterns in time. After amplifying the received signal, the TRM retransmits the time-reversed version of the signal (last in, first out). During the two-way propagation between the PS and TRM, some higher-order modes that have more interactions with an ocean bottom are removed by waveguide attenuation. Then, the remaining modes arrive simultaneously at the range of the PS and realize a focus at the original position.

The focal size or the resolution[15] of a TRM is determined by the remaining effective modes. The focal size in depth and range can be estimated roughly as

$$\Delta_Z \sim \frac{D}{M}, \tag{1}$$

$$\Delta_R \sim \frac{2\pi}{k_1 - k_M}, \tag{2}$$

where D is water depth, M is the number of effective modes, and k_1 and k_M are the horizontal wave numbers of the first and last effective modes, respectively. The vertical resolution is determined by the mode functions while the horizontal resolution is determined by the difference of wave numbers that is the smallest *interference pattern* of the acoustic field. Waveguide attenuation increases the focal size by reducing the number of effective modes. The focal size is also proportional (inversely proportional) to the wavelength (frequency). The focal sizes have been shown to approach the diffraction limit of the TRM in a static ocean.[15]

In a nonstatic medium where sound propagates through different environments each way, sound-speed fluctuations change the mode functions as well as the horizontal wave numbers. The mode functions determine the amplitude of the acoustic field while the horizontal wave numbers are associated with the phase of the wave field. Acoustic fields are very sensitive to small changes in the phase terms, indicating that the horizontal focal structure is vulnerable to sound-speed fluctuations. Changes in horizontal wave numbers cause perturbed group speeds resulting in delays or advances of mode arrivals. The detailed focal structure response to the sound-speed fluctuation is investigated in Sections III and IV.

III. TIME-REVERSAL EXPERIMENTS

In this section, we describe the measured time-reversal foci in a time-evolving ocean environment. The results show that the effect of ocean fluctuations on the time-reversal focus is not negligible in the high frequency (3.5 kHz) transmissions.

A. Experimental setup

A series of time-reversal experiments was performed in two areas near Formiche and Elba islands off the west coast of Italy depicted as shaded boxes in Fig. 2. Both areas are relatively flat environments with a water depth of about 130 m in the Formiche area and 120 m in the Elba area, respectively. A detailed oceanographic survey was accompanied by deploying thermistor chains, wave riders, current meters, etc. Sound-speed profiles (SSP) also were measured frequently by conductivity-temperature-depth (CTD) casts. The SSPs collected over the experiment period indicate variability of the upper water column, especially in the thermocline region. The downward-refracting sound speed structure resulted in a substantial interaction of the sound field with the ocean bottom. This also suggests that acoustic fluctuations are dominated by sound-speed fluctuations caused by internal waves rather than surface wave effects.

Figure 3 shows the components of a TRM experiment. The detailed hardware configurations for the low- and high-frequency TRMs can be found in Refs. 4 and 16, respectively. The TRM received an incident signal from a probe source (PS) and retransmitted the time-reversed signal to the vertical receiver array (VRA) that was collocated with the PS. In the April 1996 and May 1997 experiments, a TRM was deployed near the island of Formiche di Grosseto. The TRM spanned 77 m of the water column with 20 receiving hydrophones and 20 contiguously located slotted cylinder sources. Each element of the TRM was operated at a nominal source level of 160 dB *re* 1 μPa per transducer with a center frequency of 445 Hz. A PS was located in the vicinity of a 48-element VRA. The July 1999 experiment was performed

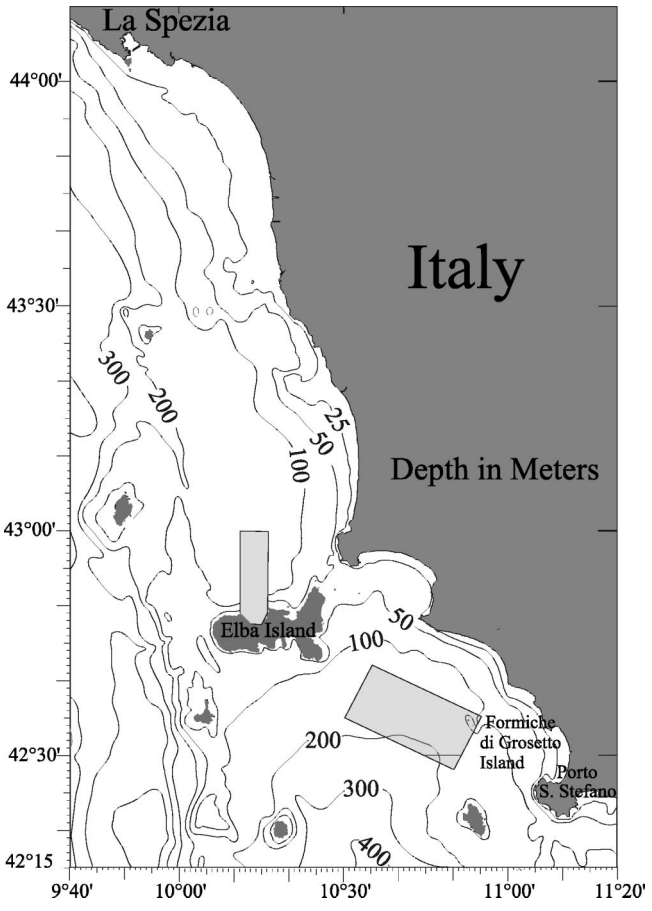


FIG. 2. Locations of the time-reversal experiments. Low-frequency (445 Hz) experiments were performed near Formiche di Grosseto and high-frequency (3.5 kHz) experiments were carried out both in the Formiche and Elba areas.

with a center frequency of 3.5 kHz in both the Formiche and Elba areas and the May/June 2000 experiment was performed in the Elba area. The experimental setup was similar to the lower-frequency experiments but the TRM had 29 transducers spanning a 78 m aperture with a nominal source level of 180 dB *re* 1 μPa per transducer. The VRA covered 93 m of the water column with 32 hydrophones.

B. Time-reversal foci in a fluctuating ocean

If the ocean is static, PS pulses can be stored ahead of time to focus at specific locations. However, the temporal variability of the ocean is expected to limit such a procedure. In the May 1997 experiment, we found that probe pulses at 445 Hz still produced a significant focus at the original probe–source location up to one week later (see Hodgkiss *et al.*[7] for details). But our recent experiment in May/June 2000 with 3.5 kHz transmissions showed that a stable focus was severely limited by the medium fluctuations.

Figure 4 shows the time-reversed pulses received with the VRA by retransmitting the received probe pulse after a short time (16 min), 1 day, and 1 week, respectively. The range between the TRM and PS was 15.2 km and the PS depth was 81 m. The PS signal was a 50 ms pulse at a 445 Hz carrier. The vertical focal size was about 20 m. During the period, the focus moved a few meters downward, resulting from the increased temperature in the upper part of the water column during the experiment period. Although the focus was degraded significantly after a week along with the appearance of a sidelobe in the upper water column, the TRM clearly retained a focus. This result reveals that time reversal at 445 Hz was robust in these ocean conditions.

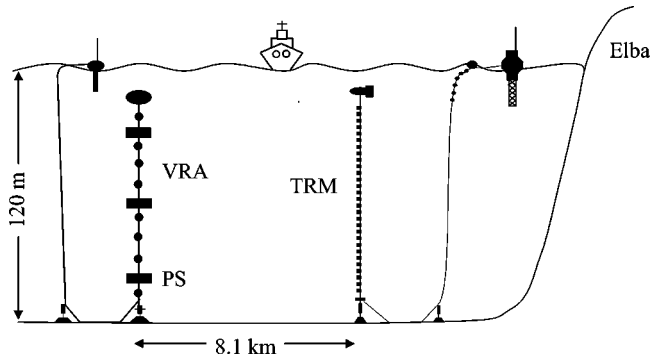


FIG. 3. Experimental setup for the high-frequency (3.5 kHz) time-reversal experiment. The time-reversal mirror (vertical source–receiver array) consisted of 29 transducers with interelement spacing 2.786 m spanning 78 m of the water column.

Figure 5 shows typical examples of a probe signal received by the TRM and the corresponding focused pulse observed by the VRA after retransmitting the received signal from the TRM. The probe source was located 7.9 km away from the TRM at a depth of 43 m. The signal was a 2 ms pulse consisting of seven cycles of a 3.5 kHz carrier. The probe signal received by the TRM shows dispersion by more than 20 ms at about 10 dB down in intensity. As expected from the measured sound-speed profile, most acoustic energy is distributed in the early arrivals, i.e., in the lower-order modes, and below the thermocline, which implies substantial interactions with the bottom. The signal above 20 m depth is negligible, indicating that the interaction with the ocean surface is not important for propagation to this distance. The focused signal shows the spatial and temporal compression obtained with the time-reversal process. The vertical focal size was about 2 m and the signal length was compressed back to 2 ms at 3 dB down points.

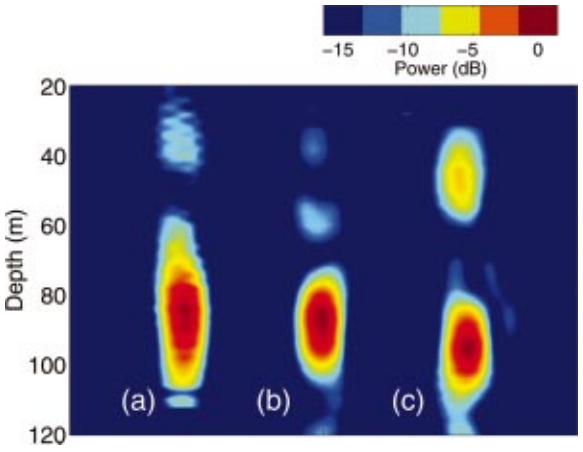


FIG. 4. Measured time-reversal foci with low-frequency (445 Hz) transmissions. A 50 ms pulse received by a TRM from a PS 15.2 km away at a depth of 81 m was retransmitted (a) 16 min, (b) 1 day, and (c) 1 week later.

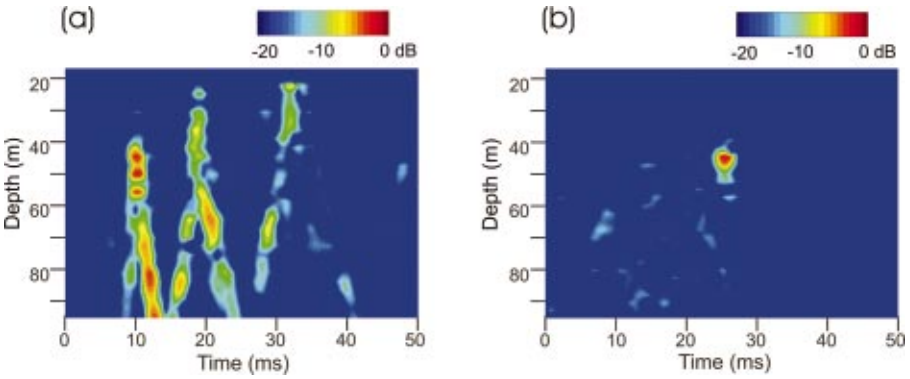


FIG. 5. Measured examples of (a) a probe signal received by the TRM and (b) a focused signal observed by the VRA. The distance between the probe source and the TRM was 7.9 km and a 2 ms pure tone pulse at 3.5 kHz was transmitted at a depth of 43 m.

We also measured the temporal variability of the focused signal with high-frequency transmissions. Figure 6(a) shows the results of high-frequency time-reversal foci obtained by retransmitting the stored PS pulse every minute over 45 min starting at 08:28 GMT on 31 May 2000. The pulse was 10 ms long and the probe source was 8.1 km away from the TRM at a depth of 68 m. The vertical focal size was less than 2 m in this case. Although the vertical focal size did not change very much, the intensity at the PS position decreases gradually to −7 dB over 45 min, as shown in Fig. 6(b). In addition to the gradual decrease in intensity with time, another feature is that the focal structure shows a slight upward movement between 20 and 30 min and a corresponding sidelobe appears at a depth of around 75 m. The intensity curve at the probe source depth indicates additional degradation of the focal intensity during this short time period. This phenomenon was more prominent in the next measurement.

Another measurement (Fig. 7) obtained a few hours later (starting at 11:04 GMT on 31 May 2000) showed more unstable results of the time-reversal foci. In this case, a 2 ms pulse was generated from the PS. The PS depth was 54 m and the range was the same (8.1 km) as before. The retransmission from the TRM began 1 min after the reception of the PS pulse and lasted for 50 min. The focal structure shows more sidelobes overall. Once again, the foci moved upward between 18 and 33 min with a sidelobe below at 60 m. The foci almost were divided into two vertical structures during this time and they then were recombined into one later. The intensity level at the probe source position as shown in Fig. 7(b) decreased rapidly by −8 dB during the time of the split focal structure then recovered up to −3 dB at 40 min.

From the experimental results we can extract some important facts about the focal structure in a time-evolving ocean environment. First, as expected, the high-frequency focal structure was much more sensitive to the environmental changes. Although the measurements were not sufficient to define quantitatively the relaxation time of the focal structure, it is reasonable to estimate that it was on the order of about a day with the 445 Hz transmissions and less than an hour with the 3.5 kHz transmissions in the ocean conditions of the area. Second, the focal intensity at the position of the PS did not decrease linearly with time. In a dynamic ocean environment, we observed focal splitting and recombination.

IV. NUMERICAL SIMULATION OF TIME-REVERSAL FOCUSING

In the previous section, we examined the variation of vertical focal structures measured with a vertical-receiver array (VRA) collocated at the position of the probe source (PS). In this section we investigate the variation of the horizontal focal structure using numerical simulations. A fundamental question is whether the focus actually disappeared or simply was somewhere else spatially during the time when we observed splitting of the focal structure. The main result of this numerical study is that a sound-speed perturbation causes a focal shift along with intensity degradation.

A. Procedure

Here, we describe the numerical simulation procedure. The first step is generating the time series of the sound-speed fluctuations. The major sources of acoustic fluctuations for the propagation of sound over the 10 km range and time scales of a few minutes to an hour are surface waves and internal waves. As we observed from the experimental results, surface wave effects are negligible in this downward-

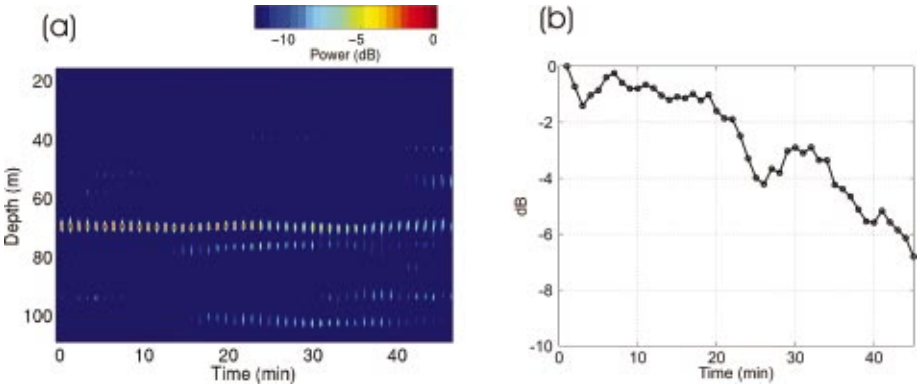


FIG. 6. Measured time-reversal stability with high-frequency (3.5 kHz) transmissions. A stored 10 ms probe pulse received by a TRM from a PS at a range of 8.1 km and a depth of 69 m was retransmitted repeatedly for 45 min; (a) the pulse signal received every minute by a vertical receiver array; (b) variation of the intensity level at the PS position.

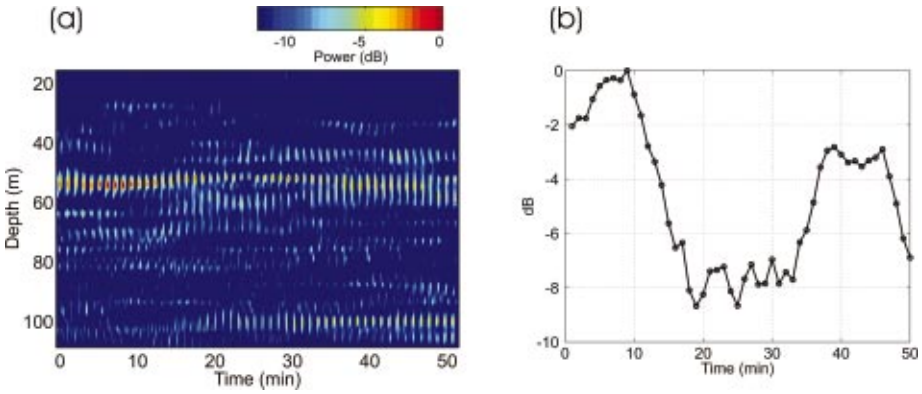


FIG. 7. The same as Fig. 6, except the probe source depth was 54 m, the pulse length was 2 ms, and the transmission lasted for 50 min.

refracting sound-speed environment. Thus, we consider only sound-speed fluctuations caused by internal waves. The sound-speed perturbation $\delta c(r,z,t)$ is assumed to be related solely with the vertical displacement of the internal waves and is given by

$$\delta c(r,z,t) = c_p[z - \eta(r,z,t)] - c_p(z), \tag{3}$$

where $c_p(z)$ is the potential sound speed and $\eta(r,z,t)$ is the vertical displacement of internal waves. The method used for the calculation of $\eta(r,z,t)$ and associated parameters are discussed in detail in the Appendix following the method developed by Henyey *et al.*[17] Refer to the listed articles[18–20] for a more thorough description of this subject. The parameters used for the generation of the internal waves were chosen to simulate the environmental condition of the May/June 2000 experiment performed near Elba Island. A representative buoyancy frequency profile in Fig. 8(b) was determined from the average value of the CTD casts obtained on 21 May 2000 near the TRM, as shown in Fig. 8(a). The maximum buoyancy frequency is about 18 cph at 25 m and exponentially decreases with depth. The power spectrum of internal waves was calculated using thermistor chain data sampled every 5 min, as shown in Fig. 8(c). An internal wave spectrum was obtained from the data set measured at 39 m. The spectrum shapes measured at the other depths were similar. The thick dashed line represents the functional form of the spectrum (see the Appendix for details) used for the internal wave simulation. The interval between the inertial frequency ω_I and the maximum buoyancy frequency $N_{\max}$ determines the range of the internal wave frequencies to be calculated. Figure 8(d) shows an example of simulated sound-speed profiles in range and depth. The fluctuations are mostly distributed between 20–60 m and become zero at the surface and bottom as specified by the boundary conditions. Two-dimensional sound-speed fields were calculated every minute for several hours for use as environmental input for modeling acoustic propagation.

Next, the acoustic field between the TRM and field points was calculated using a parabolic equation model.[21] The input sound-speed environment was range dependent, as simulated by the internal wave model. The model ocean and the array configuration were similar to the May/June 2000 experiment. The model TRM consists of 29 elements span-

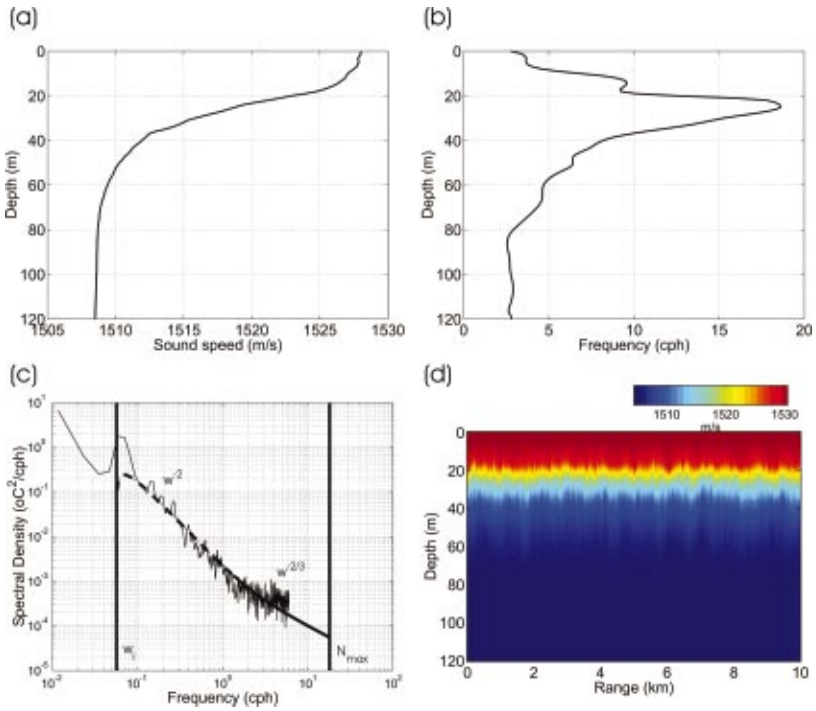


FIG. 8. Measured and simulated internal wave fields: (a) measured average sound speed profile; (b) profile of buoyancy frequency calculated from CTD data; (c) measured internal wave spectrum (solid line) and a function used for the internal wave simulation (thick dashed line); and (d) simulated sound-speed structure.

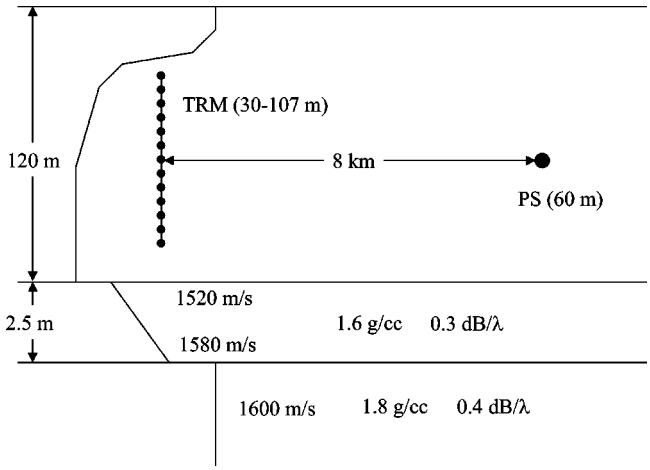


FIG. 9. Ocean waveguide parameters used for simulations of the time-reversal process. The sound speed of the water column at the bottom interface is about 1508 m/s. Thus, acoustic waves propagating above the critical angle (7.2°) are highly attenuated.

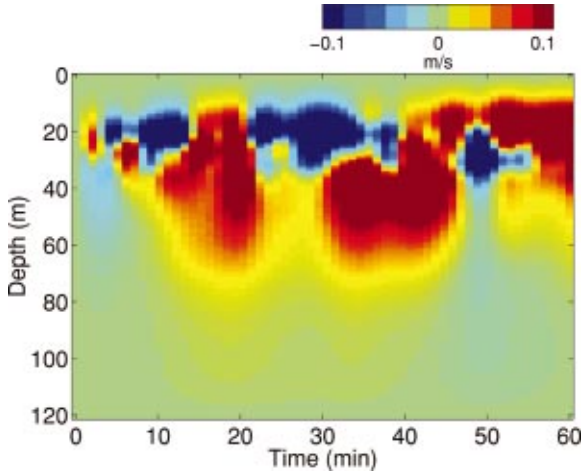


FIG. 10. Simulated sound-speed perturbations averaged in range over 8 km. The positive perturbation means increased sound speed due to the downward displacement of the internal waves.

ning the 120 m deep water column from 30 to 107 m with 2.75 m interelement spacing. The PS is located at 60 m depth and is 8 km away from the TRM. The ocean bottom has a 2.5 m thick sediment layer with a sound speed at the top interface of 1520 m/s. Since the sound speed of the water column at this depth is 1508 m/s, acoustic waves propagating above the critical angle of 7.2° are highly attenuated. The waveguide environment used for the simulation is depicted in Fig. 9. In the PE calculation, the depth and range steps were 0.1 and 1 m, respectively, and four Padé coefficients were used. The calculations were carried out at one-minute intervals for each time frame of the internal wave-induced range-dependent sound-speed fluctuations. During the propagation time between the TRM and PS (about 5 s), we assumed the ocean environment is frozen.

The final step was to implement the time-reversal process with the calculated transfer functions between the TRM and the field points. Let the Green's function at frequency ω and the jth element of the TRM due to a probe source located at a horizontal distance R and depth z_s be $G(\mathbf{r}_j;\mathbf{r}_s,\omega)$, where $\mathbf{r}_j=(0,z_j)$ and $\mathbf{r}_s=(R,z_s)$. Then, the acoustic field observed at a field point $\mathbf{r}=(r,z)$ reduces to a sum over the TRM transducer positions,

$$p(\mathbf{r},\Delta t,\omega)=\sum_j S^*(\omega)G^*(\mathbf{r}_j;\mathbf{r}_s,t_1,\omega)G(\mathbf{r};\mathbf{r}_j,t_2,\omega), \tag{4}$$

where $S(\omega)$ is the source spectrum and $[\ ]^*$ is complex conjugation. The specification of $\Delta t=t_2-t_1$ makes explicit the time delay taken for the time-reversal process between the forward probe signal propagation and the backpropagation from the TRM. As mentioned above, we assumed that environmental changes during the short time interval of the two-way propagation between the PS and TRM are negligible. The calculated time-reversed acoustic fields then were inverse Fourier transformed to obtain the pressure time series at $\mathbf{r}=(r,z)$.

B. Simulation results

Here we describe the simulation results based on the methods described above. A special emphasis is given to the relationship between the sound-speed perturbations and the changes of focal structure as well as the intensity variations at the PS position. The results provide physical insight into how the time-evolving ocean environment affects the variation of the spatial focal structure.

Figure 10 shows one hour of the simulated sound-speed perturbations. Although full range-dependent sound-speed data was used for the propagation modeling, to make the complicated three-dimensional (range, depth, and time) data understandable it is represented here as the perturbed sound speed from the sound-speed distribution at $t=0$ min and averaged over the 8 km range between the TRM and PS,

$$\delta c(z,t)=\frac{1}{R}\int_0^R [c(r,z,t)-c(r,z,0)]dr, \tag{5}$$

where R is the distance between the TRM and PS. Thus, no perturbation exists at zero minute. The positive perturbation indicates an increase in sound speed resulting from the net effect of the downward displacement of internal waves between the TRM and PS. This specific one hour realization shows two big downward internal wave displacements between 10 and 50 min. During this downward motion, the upper part of the water column also shows negative sound-speed perturbations, indicating that more than two internal wave modes are involved in this event.

Figure 11(a) shows the simulated single-frequency (3.5 kHz) focal structures at four different times (0, 10, 40, and 50 min) using the same probe signal received at 0 min. The probe signal was excited at a depth of 60 m and 8 km away from the TRM. The intensity field was normalized by the value at $t=0$ min obtained at the probe source position. The focal size at 0 min is about 2 m in depth and 200 m in range at the 3 dB down points, which is similar to the vertical focal size measured during the experiment, as shown in Figs. 5, 6, and 7. At 10 min, the downward internal wave motion starts to perturb the sound speed from the upper part of the water

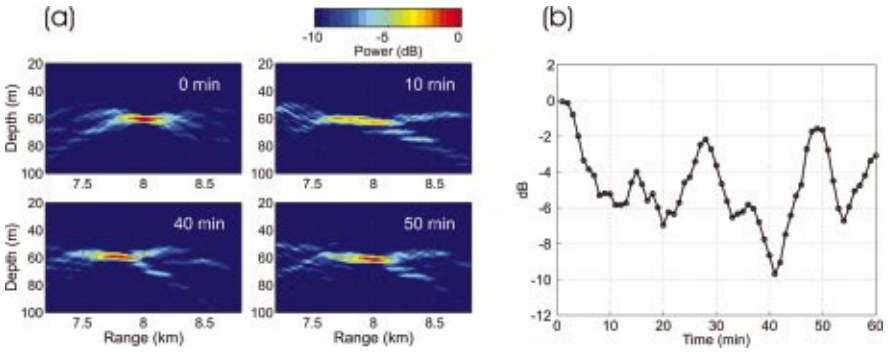


FIG. 11. Simulated single frequency time-reversal foci at 3.5 kHz. (a) Time-reversal foci at four different times (0, 10, 40, and 50 min) using the same probe signal received at 0 min from the range of 8 km and depth of 60 m. (b) The variation of the focal intensity of the time-reversal foci at the position of PS.

column (Fig. 8). Although the unperturbed lower-order modes that are trapped under the thermocline still establish a focus at the original position of the PS, the perturbed higher-order modes tend to move the focus toward the TRM, resulting in the extended focal size in range and the degradation of the focal intensity. At 40 min, a large internal wave motion causes a positive sound-speed perturbation in the upper-half part of the water column. Now all the propagating modes are perturbed in the same direction and make a focal shift toward the TRM by 200 m without much degradation of the intensity at the shifted focal center. At the PS range (8 km), the sidelobes at the end of the focus dominate the structure, which is divided into two parts as observed in the experiment. This suggests that in the experiment we were looking at the outside of the focus during the divided focal structure and the focal center was shifted in range without disappearing. As the sound-speed perturbation decreases the focus returns to the original position at 50 min. As observed in the simulation results, the changes in the vertical focal size is negligible during the one hour period while the size and position of the horizontal focal structure is quite variable.

The variation of the focal strength at the probe source position is depicted in Fig. 11(b). The intensity level is normalized by the strength at 0 min. Therefore, the negative values mean intensity degradation due to the sound-speed mismatch. The intensity curve starts at 0 dB as a maximum and runs up and down according to the degree of sound-speed perturbation. As discussed, the degradation for the first 10 min is due to the extended and diffused focal structure resulting from the sound-speed changes in the upper part of the water column. The minimum intensity level appears as a 10 dB degradation around 40 min resulting from the focal shift. The focal strength recovers up to -2 dB at 50 min when the mismatch is minimized.

The simulation results seem to have captured the characteristics of the measured data shown in Figs. 6 and 7. The main features are splitting of the focal structure associated with the large intensity drop at the corresponding time and remerging of the focus later at the correct position. The simulation results also show that the focal strength is closely related to the focal shift due to the sound-speed mismatch between the forward probe signal propagation and the backpropagation of the time-reversed signal. The intensity variation should be higher in the higher-frequency transmission for the same distance of focal shift, as described in the next section.

V. THEORETICAL ANALYSIS OF FOCAL SHIFT AND DEGRADATION

This *environmentally driven* focal shift is similar to the offset of the estimated source location in matched-field processing called a mirage.[22] In Ref. 22, the offset or error in the source localization was due to the mismatch between the model bathymetry used for the calculation of the replica fields and the real bathymetry of the experimental area. A slight frequency change between the forward and backward propagation in time reversal also provides a shift in focal range.[7,23] This technique is based on waveguide-invariant theory,[10,11,13] which explains the linear relationship between the rate of the change of the phase along the waveguide and the rate of change of the acoustic frequency. Here, we examine the focal shift associated with the frequency change as well as the environmental variation together using a generalized waveguide-invariant theory.[12] This investigation provides a clue for developing a method of robust focusing, which is discussed in the next section.

We employ the adiabatic normal mode theory[24] to describe acoustic propagation in a range-dependent ocean, assuming that the ocean environment changes gradually in range. Although mode coupling effects are not negligible for 3.5 kHz transmissions, the approach still is useful to explain the main features in the case of small environmental fluctuations and sufficient to provide physical insight into how focal shift and intensity degradation are related with various waveguide and acoustic parameters. The sound-pressure field received at a range $r=0$ and depth z due to a point source of unit strength at $r=R$ and $z=z_s$ can be written as

$$p(0,z)=\frac{\sqrt{2\pi i}}{\rho(z_s)}\sum_n \frac{u_n(R,z_s)u_n(0,z)}{\sqrt{\bar{\xi}_n R}}\exp(i\bar{\xi}_n R), \qquad (6)$$

where $i=\sqrt{-1}$, u_n is the local modal function, and $\bar{\xi}_n$ is the range-averaged horizontal wave number given by $\bar{\xi}_n = R^{-1}\int_0^R \xi_n\, dr$. Normally, horizontal wave numbers are complex and can be divided into the real and imaginary part as $\bar{\xi}_n = k_n + i\bar{\alpha}_n$.

Substituting Eq. (6) into Eq. (4), ignoring the small changes of the modal functions during the elapsed time $\Delta t = t_2 - t_1$ for the time-reversal process, and using the orthonormality property of mode functions, the time-reversed acoustic field at a field point can be written as

$$p_N(r,z,\Delta t)\sim\sum_n A_n \exp[-\bar{\alpha}_n(R+r)]$$
$$\times\exp i[\bar{k}_n(t_2)r-\bar{k}_n(t_1)R], \quad (7)$$

where

$$A_n\simeq\frac{u_n(r,z)u_n(R,z_s)}{\rho(z_s)\sqrt{\bar{\xi}_n^2 rR}}. \quad (8)$$

Here we keep the variation of horizontal wave numbers in the phase term but the variations are ignored in the modal attenuation and amplitude term A_n.

The intensity field around the PS position can be written as

$$I(r,z,\Delta t)\equiv|p_N|^2\sim\sum_{n,m} A_m A_n \exp[-\bar{\alpha}_{mn}(R+r)]$$
$$\times\cos[\Delta\bar{k}_{mn}(t_2)r-\Delta\bar{k}_{mn}(t_1)R], \quad (9)$$

where $\bar{\alpha}_{mn}=\bar{\alpha}_m+\bar{\alpha}_n$ is modal attenuation and $\Delta\bar{k}_{mn}=\bar{k}_m-\bar{k}_n$ is the real part of the wave number difference between the mode numbers m and n. A perturbed horizontal wave number difference $\delta\Delta\bar{k}_{mn}$ due to small changes in sound speed $\Delta c(r,z)$, bathymetry $\Delta h(r)$, or acoustic frequency $\Delta\omega$ during the time Δt can be expressed as

$$\Delta\bar{k}_{mn}(t_2)\simeq\Delta\bar{k}_{mn}(t_1)+\delta\Delta\bar{k}_{mn}(\Delta c,\Delta h,\Delta\omega). \quad (10)$$

At range $r=R$, the closure property of the modes, $\Sigma_n u_n(z)u_n(z_s)/\rho(z_s)\simeq\delta(z-z_s)$ and $\delta(z)$ is the Delta function, implies that the focus is at the depth $z=z_s$. However, the modal attenuation and the perturbed modal wave number difference degrade the focal strength by

$$\exp(-2\bar{\alpha}_{mn}R)\cos(\delta\Delta\bar{k}_{mn}R). \quad (11)$$

The modal attenuation degrades the focus by reducing the amplitude of the intensity field while the perturbed modal wave number difference degrades the focus by shifting the focal center to a different range since it is associated with the phase term.

According to first-order perturbation theory, the perturbed modal wave number caused by small changes in sound speed $\Delta c(r,z)$ or frequency $\Delta\omega$ can be represented as[25]

$$\delta\bar{k}_n=\frac{1}{2k_n}\int_0^\infty\frac{\delta\bar{q}|u_n(z)|^2}{\rho(z)}dz, \quad (12)$$

where $\bar{q}=(\omega/\bar{c})^2$ and $\delta\bar{q}$ is due to either sound-speed or frequency perturbations. For arbitrary sound-speed profiles, Eq. (12) can be used to calculate the phase term in Eq. (11). Analytic forms of horizontal wave numbers are known for some simple waveguide conditions.[11,24] One of them is the ideal waveguide for which the horizontal wave number is $k_m=\sqrt{(\omega/c)^2-(m\pi/h)^2}$, where c is the homogeneous sound speed and h is the water depth. The wave number difference then can be represented as

$$\Delta k_{mn}(c,h,\omega)\simeq\frac{\pi^2(n^2-m^2)}{2}ch^{-2}\omega^{-1}, \quad (13)$$

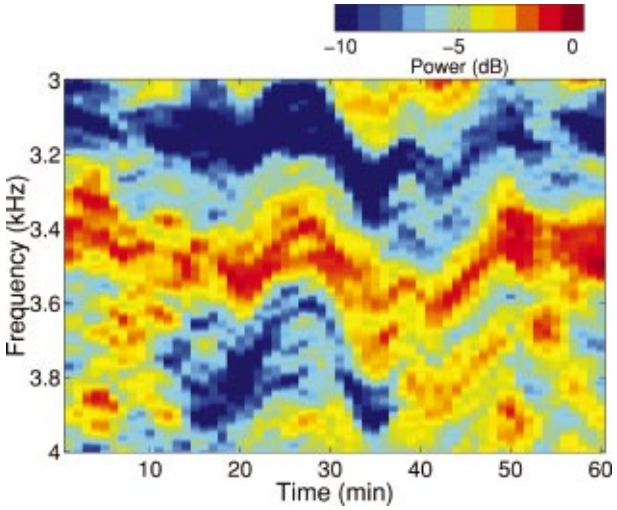


FIG. 12. Spectrogram of the simulated acoustic field of one-way propagation between the PS and a TRM element (77 m). The result shows the shift of frequency of the high-intensity zone with changing sound-speed conditions.

which is true for high-frequency propagation with a large number of modes.[11]

A measure of focal shift caused by perturbation of the wave number difference Δk_{mn} is obtained by following the stationary phase (constant intensity lines) in Eq. (9). Defining $\Phi\equiv\Delta k_{mn}r$, the stationary phase condition becomes

$$\delta\Phi=\frac{\delta\Delta\bar{k}_{mn}}{\Delta\bar{k}_{mn}}+\frac{\delta r}{r}=0. \quad (14)$$

Inserting Eq. (13) into Eq. (14) and taking derivatives with respect to h and ω, the stationary phase condition leads to the expression of waveguide-invariant theory for an ideal waveguide,[12]

$$\frac{\delta r}{r}=\frac{1}{\beta}\frac{\delta\omega}{\omega}-\frac{\gamma}{\beta}\frac{\delta h}{h}, \quad (15)$$

where $\beta=1$ and $\gamma=-2$. This expression describes the focal shift induced by a change in frequency or water depth in an ideal waveguide. The relationship between the first two terms indicates that increasing the frequency leads to increasing the focal range, which has been used for variable range focusing in the time-reversal process.[7,23] The relationship between the first term and the third term shows that increasing the channel depth also increases the focal range, which was observed as a mirage effect in matched-field processing[22] and a shift of interference patterns caused by tidal variations.[26] This expression also explains the internal-wave-induced focal shift observed in the simulation results. The internal-wave displacement changes the thermocline depth. Changing the thermocline depth has the same effect as changing the effective waveguide depth for the acoustic modes trapped below the thermocline. Equation (15) indicates that the downward internal-wave displacement (reduced channel depth) moves the focal structure toward the TRM (reduced focal range), as observed in the simulation. The distance of focal shift is independent of frequency for the same amount of channel depth (sound-speed) variation

while the acoustic fluctuation normally is greater at higher frequency due to the smaller focal size.

Equation (15) also contains information about the relationship between a change of frequency and a change of channel depth. At the fixed range R, a change in channel depth results in a change in the frequency content of the intensity field. Figure 12 is a spectrogram of the simulated intensity field generated by the one-way propagation between the PS and a receiver (77 m) of the TRM. The spectrogram shows the positive frequency shift as the internal wave moves downward, as indicated by Eq. (15). This is simply an interference phenomenon and does not imply a nonlinear process. The frequency shift caused by hydrodynamic variability including internal waves has been discussed earlier.[27–30]

In other waveguide conditions, the parameters β and γ in Eq. (15) have different values[12] for a different group of modes. Specifically for β, from the definition of the relationship between the change in horizontal wave number difference and the change in frequency,

$$\frac{\delta\Delta\bar{k}_{mn}}{\Delta k_{mn}} \equiv -\frac{1}{\beta_{nm}}\frac{\delta\omega}{\omega}, \tag{16}$$

where $\bar{\beta}_{nm}$ can be simply calculated using the modal slowness,[13]

$$\bar{\beta}_{nm} = -\frac{\Delta\bar{k}_{mn}/\omega}{\delta\Delta\bar{k}_{mn}/\delta\omega} = -\frac{\Delta\bar{s}_{p,mn}}{\Delta\bar{s}_{g,mn}}. \tag{17}$$

Here $\Delta\bar{s}_{g,mn} = \bar{s}_{g,m} - \bar{s}_{g,n}$ and $\Delta\bar{s}_{p,mn} = \bar{s}_{p,m} - \bar{s}_{p,n}$. The $\bar{s}_{g,m}$ and $\bar{s}_{p,m}$ are the range-averaged modal group slowness and modal phase slowness of mode m, respectively. When multiple mode groups are involved, focal shifting inevitably is accompanied by a focal strength degradation. However, the perturbed horizontal wave numbers between the modes within a dominant modal group can have a high correlation resulting in a shift of the focal structure back and forth around the PS, as observed in the simulation results.

VI. ADAPTIVE TIME REVERSAL

Now that we have some quantitative understanding of the nature of the focal instability, we investigate methods for robust time-reversal focusing. The methods involve modifying the received probe signal to make it more robust to the time-evolving ocean environment. Simulation results based on the suggested methods are discussed.

A. Adaptive MFP versus TRP

Given Eq. (11), a direct approach for reducing the sensitivity of the time-reversal process to sound-speed variations is minimizing the effect of $\cos(\delta\Delta\bar{k}_{mn}R)$ by modifying the received signal to a form suitable for the backpropagation environment. Without having exact *a priori* environmental information along the propagation path, one possible method is to reconstruct the received signal in a form that spans the range of possible wave-front perturbations. By backpropagating a signal representing the multiple wave fronts, we can minimize the phase perturbation and avoid the case of the extreme mismatch between forward propagation and backpropagation.

Considering the similarity between matched-field processing and time-reversal processing,[14] adaptive methods developed for robust MFP[31] can be applied to TRP. A simple approach for robust MFP is to use multiple replica-vector constraints, e.g., multiple location constraints (MLC)[32] or multiple sound-speed constraints (MSC).[33] The MLC method utilizes multiple neighboring points to construct a replica vector. The MSC method makes use of multiple sound-speed profiles obtained from several measurements or an archived database to generate the possible wave-front perturbations in a particular geographical region. These methods easily can be implemented with computer-generated replica fields. To apply these methods to TRP, we need to modify the measured probe signal rather than the replica fields since the ocean itself generates the replica fields in TRP. This requires additional measurements, e.g., multiple probe signals obtained at different locations or under different sound-speed conditions over time, which might not be possible in practice.

As discussed in the previous section, the frequency can be another choice as a parameter for multiple constraints since the waveguide invariant theory relates the frequency shift to the focal range shift, water depth, and sound speed. In addition, time-reversal focusing using multiple frequency constraints (MFC) requires only a single transmission of a broadband pulse as opposed to the MLC and MSC, which involve multiple probe signals.

B. Methods

The first step of adaptive TRP is constructing a probe signal matrix by gathering multiple probe signal vectors as

$$\mathbf{P}(\omega) = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N], \tag{18}$$

where $\mathbf{P}$ is a $J \times N$ probe signal matrix at frequency ω and each element $\mathbf{p}_n$ is a $J \times 1$ signal vector. J is the number of transducers in the TRM and N is the number of probe signal vectors received by the TRM. The vectors $\mathbf{p}_n$ have different meanings in the MLC, MSC, and MFC methods. In the MLC method, the vectors represent narrow band signals received from probe sources at different ranges:

$$\mathbf{p}_n = G(\mathbf{r}_j; \mathbf{r}_n, t_1, \omega), \tag{19}$$

where $G(\mathbf{r}_j; \mathbf{r}_n, t_1, \omega)$ is the Green's function at frequency ω and $t = t_1$ between the TRM at $\mathbf{r}_j$ and the probe source at $\mathbf{r}_n$. The known source spectrum $S(\omega)$ is removed from the received signal. In the MSC method, they are narrow band signals obtained over time, indicating that each vector represents a different sound-speed environment,

$$\mathbf{p}_n = G(\mathbf{r}_j; \mathbf{r}_s, t_n, \omega), \tag{20}$$

where $G(\mathbf{r}_j; \mathbf{r}_s, t_n, \omega)$ is the Green's function obtained at $t = t_n$ from a probe source located at $\mathbf{r}_s$. In the MFC method, each vector corresponds to a different frequency component extracted from a single broadband pulse,

$$\mathbf{p}_n = G(\mathbf{r}_j; \mathbf{r}_s, t_1, \omega_n), \tag{21}$$

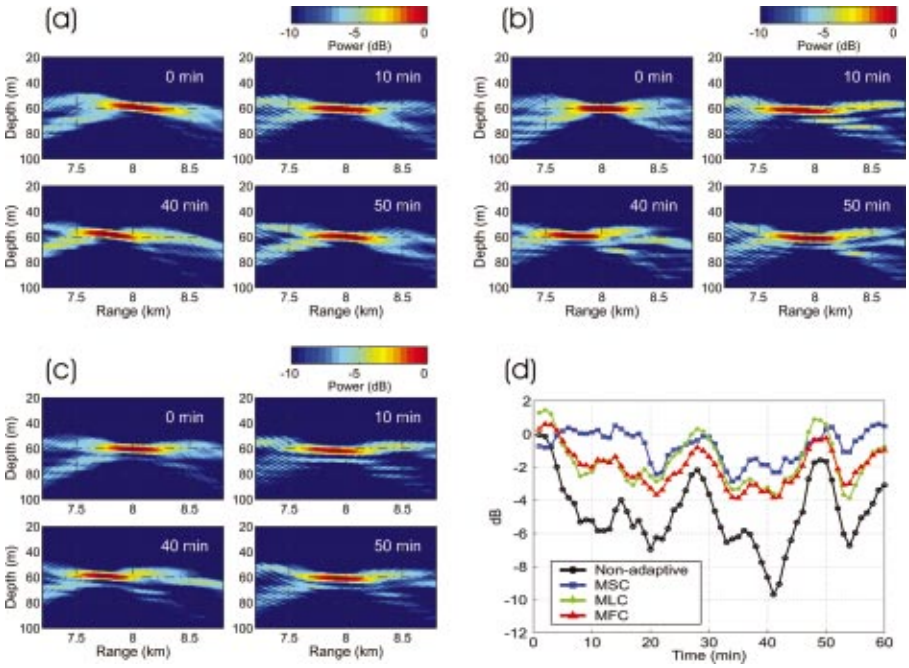


FIG. 13. Simulated single frequency time-reversal foci at 3.5 kHz using adaptive methods: (a) multiple sound-speed constraints (MSC); (b) multiple location constraints (MLC); and (c) multiple frequency constraints (MFC). Also shown in (d) is a comparison of the variation of the time-reversal focal strength at the probe source position (range at 8 km and depth at 60 m).

where ω_n is a frequency bin around ω. The performance of the adaptive methods depends on the similarity between the signal matrix constructed from the multiple constraint methods and the possible signal forms in the backpropagation environment. The signal vectors thus should be measured with enough time covering the possible oscillations of the internal waves in MSC, with ranges of the possible focal shifts in MLC around the probe source but not exceeding the focal size,[32] and with the corresponding frequency range in MFC (typically within 10% of the center frequency).

The next step is to derive a *field vector* that will be used for backpropagation. Here, we simply follow the procedure described in Ref. 33. If we take many probe signal pings or frequency bins, the number of probe signal vectors spanning all possible wave-front perturbations can be large. However, this does not mean the effective dimension of the perturbation space also is large. The modal phase perturbation can be highly correlated among the modes as well as signal pings. The dimension actually is determined by the number of effective internal wave modes interacting with acoustic modes. Normally, internal waves can be represented by a few modes in shallow water.[20] The design of an efficient constraint space for the signal vector consists of selecting the minimum number of vectors that can best approximate the phase perturbation space. This dimension reduction can be achieved using the singular value decomposition (SVD) of the signal matrix $\mathbf{P}$ with a rank K approximation,

$$\mathbf{P}(\omega) \simeq \mathbf{U}\Sigma\mathbf{V}^{+}, \tag{22}$$

where $[\,]^{+}$ is the Hermitian transpose, $\mathbf{U}$ is a $J\times K$ matrix whose columns are the left singular vectors, Σ is a $K\times K$ matrix whose diagonal elements are the singular values of $\mathbf{P}$, and $\mathbf{V}$ is a $N\times K$ matrix whose columns are the right singular vectors. Here J is the number of receivers in the TRM, and N is the number of probe-source pings or frequency bins observed. Now the field vector $\mathbf{H}$ for backpropagation can be obtained from a linear combination of the left singular vectors,

$$\mathbf{H}(\omega) = \mathbf{U}(\omega)\mathbf{q}, \tag{23}$$

where $\mathbf{q}$ is a $K\times 1$ vector representing the contributions of each singular vector. The singular values tend to decrease rapidly with an increasing number such that the first singular vector corresponding to the largest singular value can be sufficient as a field vector for stable focusing, i.e., $\mathbf{q} = [1,0,...,0]^{T}$. In other words, the effective rank of the matrix $\mathbf{P}$ is equal to one.

The final step is to replace $G(\mathbf{r}_j;\mathbf{r}_s,t_1,\omega)$ in Eq. (4) with $\mathbf{H}(\omega)$ in Eq. (23). Then the adaptive time-reversed pressure field becomes

$$p_A(\mathbf{r},t_2,\omega) = \sum_j S^*(\omega)H^*(\mathbf{r}_j,\omega)G(\mathbf{r};\mathbf{r}_j,t_2,\omega). \tag{24}$$

This approach should provide a stable focal structure since the field vector $H(\mathbf{r}_j,\omega)$ was chosen to maintain a high correlation with $G(\mathbf{r};\mathbf{r}_j,t_2,\omega)$ for all possible wave front perturbations. The above SVD process smears out the perturbation of the high-order modes that are most affected by the internal waves in a downward refracting environment. The backpropagated fields then are missing these perturbed higher-order modes resulting in the increased focal size. This has the same effect as selectively increasing the waveguide attenuation.

C. Numerical examples

We start with an example of single-frequency (3.5 kHz) time-reversal foci simulated using the proposed adaptive methods (MSC, MLC, and MFC). The field vector for each method was derived from the signal matrix $\mathbf{P}(\omega)$ constructed as follows. For the MSC method, 24 signal vectors (5 min intervals) were calculated using two hours of sound-speed data before starting backpropagation. For the MLC method,

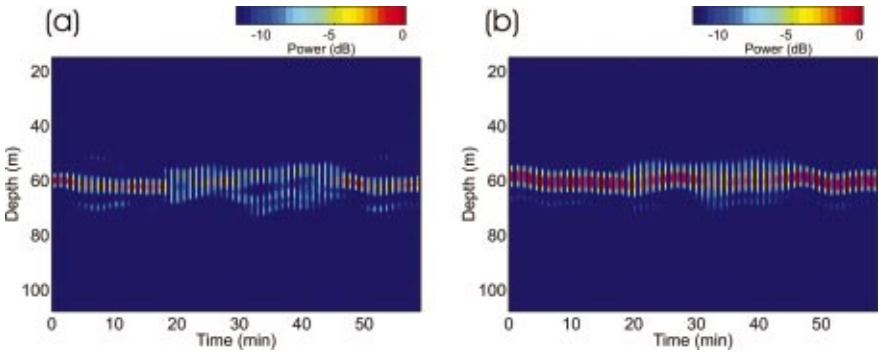


FIG. 14. Simulated 10 ms time-reversed pulse at the probe source range (8 km) (a) with the nonadaptive method and (b) with the adaptive (MFC) method.

15 signal vectors were calculated using the probe sources distributed every 20 m in range spanning 7.85–8.15 km. Last, for the MFC method, 30 frequency bins spanning 3.2–3.8 kHz were extracted from a single broadband pulse calculated with the sound-speed environment at 0 min. The derived field vectors were used to evaluate backpropagation focusing at the center frequency for the next one hour period, which was the same time period used for simulation of the nonadaptive method in Fig. 11. To compare these results with those of the nonadaptive method, the intensity field was normalized using the intensity field of the nonadaptive method as

$$I(r,z,t)=10\log\left(\frac{|p_A(r,z,t)|^2}{|p_N(r_s,z_s,0)|^2}\right), \tag{25}$$

where I is the normalized intensity field, p_A is the time-reversed pressure field with the adaptive methods, and $p_N(r_s,z_s,0)$ is the time-reversed pressure field with the nonadaptive method at $t=0$ min obtained at the probe source position.

The results (Fig. 13) clearly show that the focal size becomes larger with adaptive methods. The horizontal focal size is about 400 m and the vertical focal size is about 4–5 m at the 3 dB down points, which is twice as large as the focal size of the nonadaptive method. The focal intensity remains strong at 10 min while it showed a diffused pattern with the nonadaptive method. Although focal shifts are also observed with the adaptive methods, the increased focal size helps avoid extreme intensity drops caused by the focal shift, as seen at 40 min in the nonadaptive result. As shown in Fig. 13(d), the adaptive methods consistently show higher-intensity levels than those of the nonadaptive method. In this specific case, MSC shows better performance and the results from the other two methods (MLC and MFC) are similar to each other. As mentioned previously, the performance depends on the similarity between the field vector derived from the signal matrix and the signal forms in the backpropagation environment. In this case, the two hour period for MSC appears to be long enough to better cover the possible wavefront perturbation than the range and frequency spans used for the other two methods. At times the intensity level goes above 0 dB resulting from the increased total power in adaptive methods. In this simulation, the power of the adaptive field vectors were larger by 65% (MSC), 48% (MLC), and 40% (MFC) than that of the nonadaptive method, resulting in approximately 2 dB higher intensity. Here we used about $\pm 2\%$ range span ($\delta r/r$) for MLC for the calculation of the field vectors but had to use about $\pm 8.6\%$ frequency span ($\delta\omega/\omega$) for MFC to obtain a similar result. This indicates that the parameter β of the dominant mode group is larger than 1 in this sound-speed environment.

Last, Fig. 14 shows the simulation of a 10 ms pulse backpropagated every minute for one hour from a TRM and received by a vertical-receiver array located at the PS range. In the nonadaptive method, the received probe pulse was time reversed and transmitted without modification. In the adaptive method (MFC), the field vector was obtained using a probe signal covering the same frequency band (± 300 Hz). Since the pulse is relatively narrow band, we used the single field vector **H** obtained at the center frequency 3.5 kHz for all of the frequency bins of the 10 ms pulse. As expected, the vertical focal sizes are larger with the adaptive method. The nonadaptive method also shows a split focal structure between 20 to 45 min. The adaptive method removes this feature and maintains a single focal structure throughout the entire 60 min period.

VII. SUMMARY

We have investigated the stability of time-reversal focal structures with both experimental data and numerical simulations. As expected, the high-frequency TRM (3.5 kHz) was sensitive to the fluctuating ocean environment such that a stable focus was limited to on the order of few tens of minutes, which is in contrast to several hours to days at 450 Hz. Numerical simulations involving sound speed fluctuations due to internal waves demonstrated that the focal intensity does not decrease linearly with time but shows focal splitting and recombination associated with a focal shift caused by phase changes in the acoustic field. Using waveguide-invariant theory, the focal range shift was related to the frequency shift and channel depth changes.

A qualitative and quantitative understanding of the nature of focal stability provided guidance for developing robust time reversal focusing methods in fluctuating ocean environments. The basic idea is to find an adaptive weight vector representing all possible perturbations using the SVD prior to backpropagation. Among the three different multiple constraint methods (MLC, MSC, and MFC) considered, numerical simulations showed that the MFC is the most efficient method to extend the period of stable focusing since it requires just a single transmission of a broadband probe signal.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research, Grant No. N00014-01-D-0043-D06.

APPENDIX: INTERNAL WAVE SIMULATION

Here we briefly describe the internal wave model following the approach developed by Henyey *et al.*[17] with a little modification in the internal-wave energy spectrum. The vertical displacement $\eta(r,z,t)$ of an internal wave can be expressed as a sum of discrete normal modes as

$$\eta(r,z,t)=\int_0^{2\pi}\int_{\omega_I}^{N_{\max}}\sum_l f(\omega,l,\theta)\phi_l(\omega,z) \times\exp i(k_h r\cos\theta-\omega t)d\omega\, d\theta, \tag{A1}$$

where ω_I is the local inertial frequency, $N_{\max}$ is the maximum buoyancy frequency, and k_h and ϕ_l are the horizontal wavenumber of the lth mode and the vertical mode function, respectively. The displacement field $\eta(r,z)$ at a given time is a two-dimension cut of a three-dimensional internal wave field propagating from all azimuthal directions. The dispersion relation $\omega(k_h)$ and eigenfunctions $\phi_l(\omega,z)$ are determined by the second-order differential equation,

$$\frac{d^2\phi_l}{dz^2}+k_h^2\left[\frac{N^2(z)-\omega^2}{\omega^2-\omega_I^2}\right]\phi_l=0, \tag{A2}$$

subject to the boundary conditions $\phi_l(0)=\phi_l(D)=0$, where D is water depth, $N(z)$ is the buoyancy frequency. $F(\omega,l)=\int_0^{2\pi}f(\omega,l,\theta)d\theta$ is a zero mean, complex Gaussian random variable associated with the internal-wave power spectrum,[18]

$$|F(\omega,l)|^2=bE_{\mathrm{GM}}\rho\left(\int N(z)dz\right)^2 H(l)B(\omega), \tag{A3}$$

where bE_{GM} is the internal-wave energy parameter, ρ is the water density and $H(l)$ is the mode amplitude spectrum given by

$$H(l)=M(l^2+l_*^2)^{-p/2}, \tag{A4}$$

and M is a normalization factor such that $\Sigma_{l=1}^{l_{\max}}H(l)=1$. The p and l_* determine the relative mode amplitude by power law. We used the following functional form for the frequency spectrum $B(\omega)$ which represents the measured spectrum well as shown in Fig. 8(c):

$$B(\omega)=W\frac{\omega_I\sqrt{\omega^2-\omega_I^2}\sqrt{N_{1/3}^2-\omega_I^2+\omega^2}}{\omega^3}, \tag{A5}$$

where W also is a normalization factor such that $\int_{\omega_I}^{N_{\max}}B(\omega)d\omega=1$. The spectrum decreases with a slope of ω^{-2} and then $\omega^{-2/3}$ from $N_{1/3}=\int N(z)dz/3D$.

The parameters used for the internal wave calculations are $bE_{\mathrm{GM}}=0.142$ m, $p=4$, $l_*=1$, and $l_{\max}=15$. The shortest and longest internal wave wavelengths used for the calculations are 20 m and 50 km. The 50 frequency components were calculated on a linear scale between the corresponding wave numbers, 0.000 126 and 0.3142 rad/m. We assumed that the internal wave field is horizontally isotropic. Each directional component was calculated for 2° intervals and integrated over all directions.

[1] A. Parvulescu and C. S. Clay, "Reproducibility of signal transmissions in the ocean," Radio Electron. Eng. **29**, 223–228 (1965).
[2] M. Fink, C. Prada, F. Wu, and D. Cassereau, "Self focusing with time reversal mirror in inhomogeneous media," *Proceedings of the IEEE Ultrasonics Symposium*, 1989, Montreal, 1989, pp. 681–686.
[3] D. R. Jackson and D. R. Dowling, "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**, 171–181 (1991).
[4] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **102**, 25–40 (1998).
[5] D. R. Dowling and D. R. Jackson, "Narrow-band performance of phase-conjugate arrays in dynamic random media," J. Acoust. Soc. Am. **91**, 3257–3277 (1992).
[6] M. Dungan and D. R. Dowling, "Computed narrow-band time-reversing array retrofocusing in a dynamic shallow ocean," J. Acoust. Soc. Am. **107**, 3101–3112 (2000).
[7] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, "A long range and variable focus phase conjugation experiment in shallow water," J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[8] G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, "An initial demonstration of underwater acoustic communication using time reversal," IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[9] H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and P. Guerrini, "Demonstration of a high frequency acoustic barrier with a time reversal mirror," IEEE J. Ocean. Eng. (in press).
[10] S. D. Chuprov, "Interference structure of a sound field in a layered ocean," in *Acoustics of the Ocean*, edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982).
[11] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics* (Springer, Verlag, Berlin, 1991).
[12] G. A. Grachev, "Theory of acoustic field invariants in layered waveguide," Acoust. Phys. **39**, 33–35 (1993).
[13] G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[14] W. A. Kuperman and D. R. Jackson, "Ocean acoustics, matched-field processing and phase conjugation," in *Imaging of Complex Media With Acoustic and Seismic Waves in Topics of Applied Physics*, edited by X. Fink *et al.* (Springer-Verlag, Berlin, 2002), Vol. 84, pp. 43–97.
[15] S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Spatial resolution of time-reversal arrays in shallow water," J. Acoust. Soc. Am. **110**, 820–829 (2001).
[16] W. S. Hodgkiss, J. D. Skinner, G. E. Edmonds, R. A. Harriss, and D. E. Ensberg, "A high frequency phase conjugation array," in *Proceedings of Oceans 2001*, 2001, pp. 1581–1585.
[17] F. S. Henyey, D. Rouseff, J. M. Grochocinski, S. A. Reynolds, K. L. Williams, and T. E. Ewart, "Effects of internal waves and turbulence on a horizontal aperture sonar," IEEE J. Ocean. Eng. **22**, 270–280 (1997).
[18] S. Flatte, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).
[19] J. A. Colosi and M. G. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[20] T. C. Yang and K. Yoo, "Internal wave spectrum in shallow water: Measurement and comparison with the Garrett–Munk model," IEEE J. Ocean. Eng. **24**, 333–345 (1999).
[21] M. D. Collins, "Generalization of the split-step pade solution," J. Acoust. Soc. Am. **96**, 382–385 (1994).
[22] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265 (1999).
[23] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[24] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994).
[25] C. T. Tindle, L. M. O'Driscoll, and C. J. Higham, "Coupled mode pertur-

bation theory of range dependence,'' J. Acoust. Soc. Am. **108**, 76–83 (2000).
[26] D. E. Weston and K. J. Stevens, ''Interference of wide-band sound in shallow water,'' J. Sound Vib. **21**, 57–64 (1972).
[27] V. M. Kuz'kin, ''The effect of variability of ocean stratification on a sound field interference structure,'' Acoust. Phys. **41**, 300–301 (1995).
[28] V. M. Kuz'kin, A. V. Ogurtsov, and V. G. Petnikov, ''The effect of hydrodynamic variability on frequency shifts of the interference pattern of a sound field in a shallow water sea,'' Acoust. Phys. **44**, 77–82 (1998).
[29] V. M. Kuz'kin, ''Frequency shifts on the sound field interference pattern in a shallow water,'' Acoust. Phys. **45**, 224–229 (1998).
[30] V. G. Petnikov and V. M. Kuz'kin, *Shallow Water Variability and Its Manifestation in Interference Pattern of Sound Field*, edited by W. A. Kuperman and G. L. D'Spain (American Institute of Physics, Melville, NY, 2000).
[31] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, ''An overview of matched field methods in ocean acoustics,'' IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[32] H. 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).
[33] J. L. Krolik, ''Matched-field minimum variance beamforming in a random ocean channel,'' J. Acoust. Soc. Am. **92**, 1408–1419 (1992).

A guided wave technique for the characterization of highly attenuative viscoelastic materials

F. Simonetti and P. Cawley[a)]
Department of Mechanical Engineering, Imperial College, London SW7 2AZ, United Kingdom

(Received 24 January 2003; revised 20 March 2003; accepted 24 March 2003)

The measurement of the acoustic properties of highly attenuative materials such as bitumen is very difficult. One possibility is to use measurements of the extent to which filling a cylindrical waveguide with the material affects the dispersion relationship of the cylinder. Torsional modes have been excited using piezoelectric transducers placed at one end of the cylinder, while the phase velocity and attenuation spectra have been measured by means of laser scanning. At each frequency, under the hypothesis of linear viscoelasticity, the phase velocity and attenuation of the fundamental torsional mode have been calculated as a function of the bulk shear velocity and the bulk shear attenuation of the inner core at that frequency. The resulting phase velocity and guided wave attenuation contour plots have been employed for deriving the unknown shear properties from the measured velocity and attenuation of the guided wave. The monochromaticity of the approach has not required a particular frequency dependence of the material properties to be assumed. Results for bitumen are given. [DOI: 10.1121/1.1575749]

PACS numbers: 43.35.Mr, 43.40.Le [YHB]

I. INTRODUCTION

Viscoelastic materials are employed in a number of technical areas. For instance, in the automotive and aerospace sectors lossy materials are widely used as either adhesives or composite matrix resins. Moreover, in the oil, gas, chemical, and petro-chemical industries highly attenuative materials, such as bitumen, are often used as coatings in order to protect pipe networks from corrosion. As a consequence, to perform ultrasonic nondestructive testing of these systems[1] the acoustic properties of the viscoelastic components need to be determined accurately. For this purpose several techniques are available. Reflectometry, for example, uses the successive echoes from both the front and the back faces of a sample rigidly bonded to a delay line, the amount of the reflected energy and its phase being dependent on the sample impedance.[2,3] However, although the method is relatively easy to carry out, the measurement of the reflected signal is affected by several factors, including beam spreading, transducer coupling, and magnification of bias and variance errors in the time trace of the consecutive echoes emitted from the back of the specimen.[4] This problem can be overcome by considering the front reflection from the delay line-sample interface only. However, since the front reflection is almost insensitive to the material attenuation in the sample, only the bulk velocity can be measured. Transmission methods[5,6] are affected by the same limitations as reflectometry. However, an improvement in the signal loss is achieved as the signals have to travel only once through the thickness. The amplitude spectrum method[7,8] and the phase spectrum method[9] are often used as signal processing methods. They do not significantly change the basic limitations.

The use of guided waves is an extremely attractive alternative to solve many of the above-mentioned problems. Horizontal shear waves in strips[10] or torsional waves in rods submerged in a Newtonian liquid of known density have been employed to measure the viscosity of the liquid.[11,12] The viscosity can be obtained by measuring the guided wave attenuation since, at a given frequency, there is a unique relationship between the guided wave attenuation and the viscosity. The method is extremely rapid and ideal for measurement on line and in real time. On the other hand, these techniques cannot be used for viscoelastic materials as the elasticity and the damping of the material cannot be linked uniquely to the guided wave attenuation. In other words, the link between the guided wave attenuation and the two acoustic properties (sound velocity and bulk attenuation) provides one equation only, which does not allow the two unknown acoustic properties to be found.

In the case of a rod embedded in another medium, the phase velocity of guided waves in the rod is only very weakly dependent on the properties of the second medium (unless the two media are similar). However, with other geometries, the phase velocity in the waveguide is sensitive to the properties of the second medium. The approach presented in this paper is based on the dispersion characteristics of guided waves propagating in a hollow cylindrical waveguide filled with the unknown viscoelastic material. Since the attenuation and the velocity of the guided waves are related to the acoustic properties of the inner core, by measuring the dispersion of these waves the properties of the viscoelastic material can be determined. The first important feature of the method is that the guided wave attenuation is only a consequence of the material damping within the inner core. If the core was perfectly elastic no guided wave attenuation would occur (assuming the waveguide to be elastic), in contrast with the case of a rod embedded in an elastic space.[13] Moreover, excellent geometry control can be achieved, since the waveguide works as a mould. As laser interferometry can be used for the detection of the propagat-

a)Electronic mail: p.cawley@imperial.ac.uk

 0001-4966/2003/114(1)/158/8/$19.00

ing modes, transducer coupling problems can be avoided.

The viscoelastic material is modeled by considering the linear viscoelastic approach as shown in Sec. II (the hypothesis of linear viscoelasticity leads to a linear form of the equations of motion in the frequency domain). Moreover, no assumptions about the frequency dependence of the acoustic properties such as Kelvin–Voigt, Maxwell, or Newtonian models are made. The technique to obtain the acoustic properties of the viscoelastic core is explained in Sec. III. Bitumen properties (Sec. V) have been measured by employing the setup described in Sec. IV.

II. THEORY

A. Wave propagation in viscoelastic media

Stress wave propagation within viscoelastic materials is accompanied by the dissipation of energy due to internal loss mechanisms such as bond breakage and bond formation. The hysteretic nature of these internal processes can be taken into account by considering a linear viscoelastic model. This approach is widely employed to describe the behavior of polymers, and a number of textbooks have been devoted to this subject (see, for example, Haddad[14]).

The Fourier transformed wave equation describing the propagation of distortional waves in a homogeneous and isotropic viscoelastic medium is[15]

$$a_s^2 \nabla^2 \boldsymbol{\psi} + i\omega \boldsymbol{\psi} = 0, \tag{1}$$

where ∇ is the three-dimensional differential operator, ω the angular frequency, and $\boldsymbol{\psi}$ the equivoluminal vector potential which is related to the Fourier transformed displacement, $\mathbf{u}$, according to

$$\mathbf{u} = \nabla \times \boldsymbol{\psi}. \tag{2}$$

The frequency-dependent complex velocity a_s is given by

$$a_s^2 = \frac{\tilde{\mu}}{\rho}, \tag{3}$$

where ρ is the density, and $\tilde{\mu}$ is the complex shear modulus which depends on the shear relaxation function of the material $\mu(t)$ and the frequency

$$\tilde{\mu} = \mu_\infty + i\omega \int_0^\infty \mu(t) e^{i\omega t} dt, \tag{4}$$

where μ_∞ is the asymptotic value of the relaxation curve.

For an unbounded space the solution to Eq. (1) can be thought of as a superposition of shear harmonic waves of the form

$$\boldsymbol{\psi} = A e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{5}$$

where A is an arbitrary complex constant, $\mathbf{x}$ is the position vector, and $\mathbf{k}$ is the complex wave number vector, which satisfies the secular equation

$$\mathbf{k}\cdot\mathbf{k} = \frac{\omega^2}{a_s^2}. \tag{6}$$

The harmonic waves propagate in the direction of the real part of $\mathbf{k}$ at a phase velocity C_{ph} which depends on the modulus of the real part of $\mathbf{k}$ according to

$$C_{\mathrm{ph}} = \frac{\omega}{|\mathrm{Re}(\mathbf{k})|}. \tag{7}$$

The modulus sign is necessary in Eq. (7) because the real part of $\mathbf{k}$ is itself a vector since the propagation can be in any direction in the (x_1, x_2, x_3) space. Along the direction of the imaginary part of $\mathbf{k}$ an exponential decay with distance occurs. In general the imaginary and real parts of $\mathbf{k}$ are not parallel (for example in the case of inhomogeneous waves[16]) and these are not supposed to be parallel in the predictions of this paper. However, the acoustic properties of a viscoelastic material are defined by considering the case in which the real and the imaginary parts of $\mathbf{k}$ are parallel. In this case the phase velocity is referred to as the bulk velocity, c_s, while the modulus of the imaginary part ($|\mathrm{Im}(\mathbf{k})|$) is the bulk attenuation in nepers per unit length. The complex velocity (3) depends on the acoustic properties according to

$$a_s = \frac{c_s}{1 + i\dfrac{\alpha_s}{2\pi}}, \tag{8}$$

where α_s is the bulk attenuation in nepers per wavelength (Np/wl) and corresponds to the ratio $|\mathrm{Im}(\mathbf{k})|/|\mathrm{Re}(\mathbf{k})|$. From Eq. (8) it follows that the bulk properties of the material are frequency dependent. Moreover, these properties can be linked to the relaxation function through Eqs. (8), (3), and (4).

B. Torsional modes

Let us consider a hollow, elastic and isotropic cylinder filled with the unknown viscoelastic material. The geometry of the system suggests that cylindrical coordinates (r, θ, z) are appropriate here (r, θ, z, represent the radial, angular, and axial positions, respectively). For torsional waves the vector potential, $\boldsymbol{\psi}$, is parallel to the cylinder axis and satisfies Eq. (1). The generic solution for a cylindrical layer will be of the form

$$\boldsymbol{\psi} = [A_0 J_0(k_r r) + B_0 Y_0(k_r r)] e^{-\zeta z} e^{i\xi z}, \tag{9}$$

where A_0 and B_0 are arbitrary constants, k_r and $\xi + i\zeta$ are the projections of $\mathbf{k}$ along $\hat{\mathbf{r}}$ and $\hat{\mathbf{z}}$, respectively. Moreover, ζ represents the guided wave attenuation whereas ξ is the component of the wave number which is related to the phase velocity according to

$$C_{\mathrm{ph}} = \frac{\omega}{\xi}. \tag{10}$$

J_0 and Y_0 are the Bessel functions of the first and second kind. While for the outer tube both of the Bessel functions are present, in the inner core only J_0 exists due to the singular behavior of Y_0 at the origin.

By imposing the condition (6) on the wave number vector $\mathbf{k}$, and by employing a matrix technique based on satisfying the continuity of the stresses and displacements at the

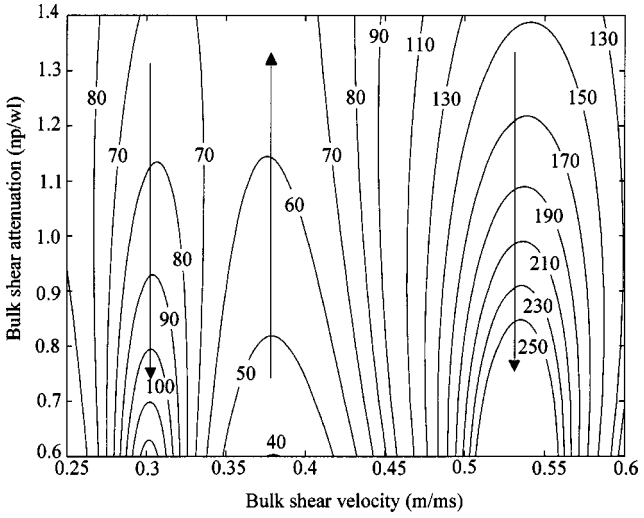


FIG. 1. Contour plot of the guided wave attenuation (dB/m) as function of the bulk shear velocity and the bulk attenuation at 50 kHz for a 6.8 mm inner radius, 0.7 mm wall thickness, filled copper tube. Core density 970 kg/m^3. The arrows indicate the direction of increasing attenuation.

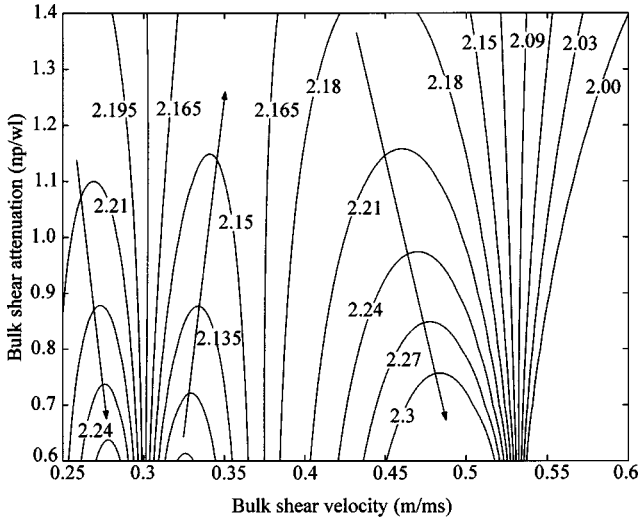


FIG. 2. Contour plot of the phase velocity (m/ms) as function of the bulk shear velocity and the bulk attenuation at 50 kHz for a 6.8 mm inner radius, 0.7 mm wall thickness, filled copper tube. Core density 970 kg/m^3. The arrows indicate the direction of increasing phase velocity.

interface between the inner core and the external tube, and the stress free condition over the external surface of the elastic cylinder, the dispersion relationship can be obtained (see, for instance, Lowe[17]). The solution to the dispersion equation at different frequencies has been calculated by using the modeling software DISPERSE[18] and is shown later in Sec. VII where the phase velocity and guided wave attenuation spectra are provided.

III. METHOD

For a given frequency and assigned tube geometry and core density, the guided wave attenuation and phase velocity are functions of the bulk shear velocity and the bulk attenuation only. Figure 1 shows the contour plot of the attenuation when all the parameters are kept constant except the acoustic properties of the inner core. Each curve corresponds to a value of ζ and provides all the values of α_s and c_s which, at the prescribed frequency and geometry, result in the specified value of guided wave attenuation. In other words, each curve provides α_s as a function of c_s and ζ,

$$\alpha_s = g_1(c_s, \zeta). \tag{11}$$

While all the couples (α_s, c_s) provided by Eq. (11) result in the same value of ζ, they do not result in the same value of the phase velocity. A second relationship can be obtained from the contour plot of the phase velocity (Fig. 2). It is possible to obtain a function that, for a given value of the phase velocity, links α_s to c_s and C_{ph},

$$\alpha_s = g_2(c_s, C_{\text{ph}}). \tag{12}$$

In this case all the couples (α_s, c_s) give the same phase velocity but different guided wave attenuation.

At a given frequency, f_0, the phase velocity, C_{ph0}, and the guided wave attenuation, ζ_0, can be measured experimentally. As a consequence, the curve $g_1(c_s, \zeta_0)$ of the guided wave contour plot and the curve $g_2(c_s, C_{\text{ph0}})$ of the phase velocity contour plot are known. By overlapping the two curves, the intersection provides the bulk acoustic properties α_s and c_s of the viscoelastic material at the prescribed frequency, f_0. This comes from the fact that at the intersection, α_s and c_s result in the measured value of the guided wave attenuation, ζ_0, as the intersection point belongs to $g_1(c_s, \zeta_0)$, and in the measured value, C_{ph0}, of the phase velocity as the intersection point belongs to $g_2(c_s, C_{\text{ph0}})$. Since the previous procedure can be applied to any frequency, it follows that the dispersion curves of the viscoelastic material $[\alpha_s(f)$ and $c_s(f)]$ can be determined over the frequency range of interest. Moreover, by contrast with other techniques,[19,20] no assumptions are made about the frequency dependence of the acoustic properties, as the method considers each frequency separately.

IV. EXPERIMENTS

The experiments were performed at the temperature of 298 K on a 1 m length copper tube (internal radius 6.8 mm, wall thickness 0.7 mm) filled with bitumen TML 24515 45/60 supplied by Shell Global Solutions. The torsional mode was excited by means of two piezo-electric transducers clamped to the external surface of the tube at one end as shown in Fig. 3. The transducers which are similar to those used in pipe testing[1] comprised shear elements mounted on a steel backing mass and oriented as shown in Fig. 3 to induce

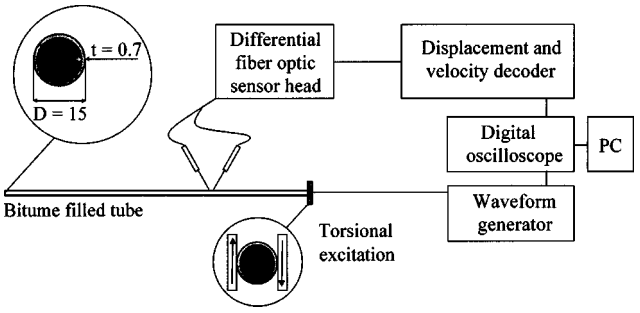


FIG. 3. Schematic diagram of the setup.

torsion. The transducers were excited by a Hanning windowed toneburst generated by a custom-made wave form generator-power amplifier. The torsional mode was detected by a laser interferometer (Sensor Head: Polytec OFV 512, Controller: Polytec OFV 3001) operating in differential mode. The tangential displacements were measured by focusing the two beams at $\pm 30°$ with respect to the radial direction and by orienting the beams in the plane perpendicular to the axis of the tube. The displacements were sampled along the tube axis and stored in a PC after 100 averages.

$T(0,1)$ is the only axisymmetric propagating mode with significant circumferential displacement below the $T(0,2)$ cutoff frequency. Therefore, only this mode will be generated below the cutoff frequency provided the excitation is symmetric. (Strictly, in attenuative systems, all the modes are propagating at any frequency.[21] However, the attenuation of the modes below their cutoff frequencies in the corresponding undamped system is very high and they are unlikely to be seen.) The use of two balanced transducers is sufficient to suppress the excitation of the $F(1,1)$ mode which is the only flexural mode in the frequency range of interest here.

Since the main goal of this paper is the measurement of the properties of the unknown material over a wide range of frequencies, a relatively wide band signal with a number of cycles ranging from three, for low frequencies, up to five for the higher frequencies was excited. More cycles are needed at higher frequencies since the attenuation tends to be higher and the use of more cycles improves the signal-to-noise ratio and reduces the effects of dispersion.

In order to obtain the dispersion curves, one can observe that for a propagating mode the Fourier transform of the signal at an arbitrary position, z, can be expressed as a function of the Fourier transform at the origin $z=0$ multiplied by a suitable complex exponential

$$U(z,\omega)=U(0,\omega)e^{-\zeta z}e^{i(\omega/C_{\mathrm{ph}})z}, \tag{13}$$

where the argument of the exponential is the projection of the complex wave number along the propagation direction. Furthermore, if the ratio R is defined as

$$R=\frac{U(z,\omega)}{U(0,\omega)}, \tag{14}$$

it follows that

$$\|R\|=e^{-\zeta z}, \tag{15}$$

$$\frac{R}{\|R\|}=e^{i(\omega/C_{\mathrm{ph}})z}=\cos\left(\frac{\omega}{C_{\mathrm{ph}}}z\right)+i\ \sin\left(\frac{\omega}{C_{\mathrm{ph}}}z\right), \tag{16}$$

where $\|\cdot\|$ is the norm operator in the complex domain. Since the ratio R can be measured experimentally, the guided wave attenuation can be determined through a linear interpolation of the experimental, logarithmic distribution of $\|R\|$ versus the axial position. Such a distribution is shown in Fig. 4. The slope of the interpolation lines gives the guided wave attenuation. Figure 5 shows the distribution of the real part of the ratio $R/\|R\|$ against the axial position. The phase velocity is computed from a cosine interpolation of the experimental data. The argument of the cosine function, ω/C_{ph}, is obtained by minimizing the square root of the error between the experimental distribution and the cosine function. This approach is particularly robust as it does not require the inversion of the cosine function to determine the unknown argument ω/C_{ph} and the successive phase unwrapping procedure due to the periodicity of the cosine function.[22]

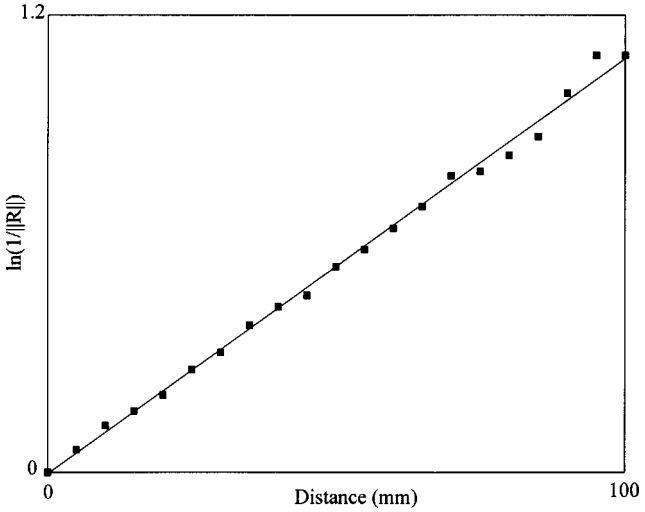


FIG. 4. $\ln(1/\|R\|)$ against the axial position z, at 50 kHz for a 6.8 mm inner radius, 0.7 mm wall thickness, copper tube filled with bitumen: (■) experiments; (—) linear interpolation.

V. BITUMEN DISPERSION CURVES

Here the procedure described in Sec. IV is applied to the characterization of the bitumen TML 24515 45/60 acoustic properties. For this type of bitumen the measured density was 970 kg/m^3.

Figure 6 shows the measured guided wave attenuation as a function of frequency. The curve is obtained as a superposition of measurements performed at different center frequencies and successively postprocessed according to the method described in Sec. IV. Figure 7 shows the experimen-

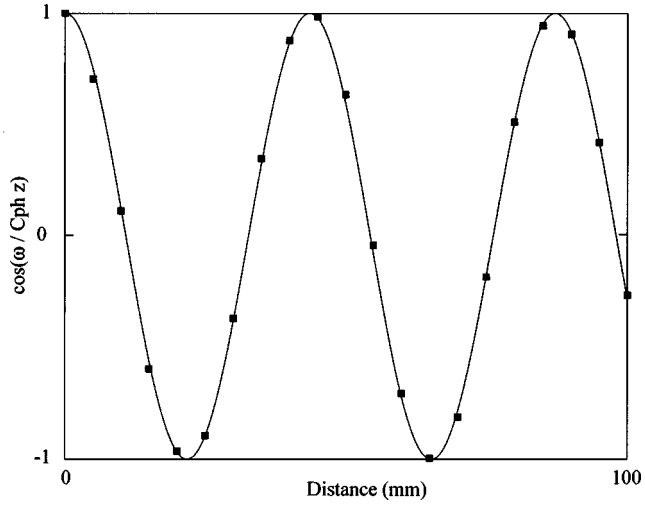


FIG. 5. $\cos((\omega/C_{\mathrm{ph}})z)$ against the axial position z, at 50 kHz for a 6.8 mm inner radius, 0.7 mm wall thickness copper tube filled with bitumen: (■) experiments; (—) cosine interpolation.

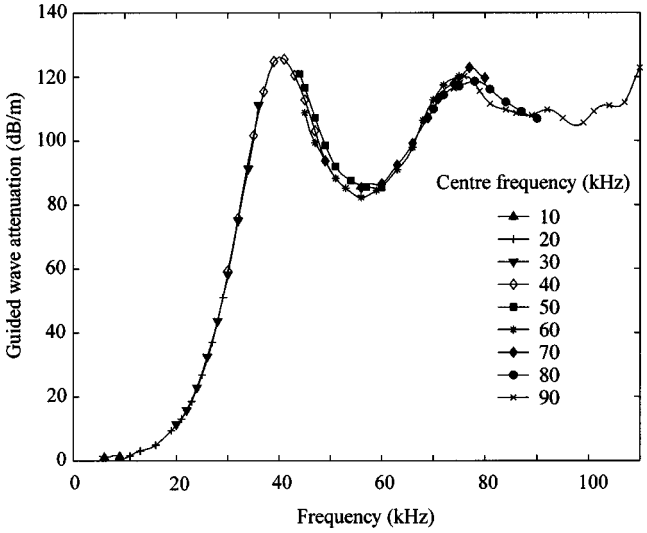


FIG. 6. Experimental guided wave attenuation versus frequency for a 6.8 mm inner radius, 0.7 mm wall thickness copper tube filled with bitumen.

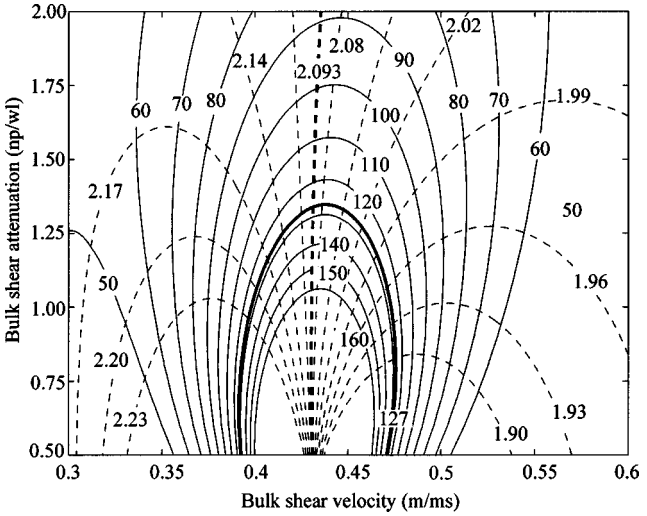


FIG. 8. Contour plot for the guided wave attenuation (solid lines dB/m) and the phase velocity (dashed lines m/ms) calculated at 40 kHz for a 6.8 mm inner radius, 0.7 mm wall thickness copper tube filled with bitumen (density 970 kg/m^3).

tal phase velocity for the filled tube. The segments of the curves obtained at different center frequencies overlap very well.

As an example, let us consider the frequency where the first attenuation peak occurs (i.e., 40 kHz, Fig. 6). At this frequency the guided wave attenuation is 127 dB/m while the phase velocity is 2.093 m/ms (Fig. 7). Figure 8 shows the overlapping of the guided wave attenuation and phase velocity contour plots at 40 kHz. The intersection of the curve at constant attenuation (ζ=127 dB/m) and the curve at constant phase velocity (C_{ph}=2.093 m/ms) provides the values of the shear velocity, 0.43 m/ms, and the bulk attenuation, 1.35 Np/wl, of bitumen at 40 kHz. It is interesting to note that the two curves intersect each other almost perpendicularly, which implies that the method is numerically robust. The bitumen bulk shear velocity versus frequency is shown in Fig. 9(a), and the bulk shear attenuation against frequency is shown in Fig. 9(b). The curves have been derived by applying the former procedure to several frequencies (black squares). The trend lines (solid lines) have been obtained by best fitting second-order polynomials. It has to be emphasized that a polynomial interpolation is not strictly representative of the behavior of a viscoelastic material as it cannot account for the transition between the material rubbery and glassy behavior. More suitable mathematical expressions are provided by Hartmann *et al.*,[23] who show how the Havriliak–Negami model can accurately describe the trend of both modulus and loss factor of a polymer over the entire frequency range. According to the Havriliak–Negami model the complex modulus is given by

$$\tilde{\mu}=\mu_{\infty}+\frac{\mu_{\infty}-\mu_0}{(1+(if/f_0)^{\alpha})^{\beta}}, \tag{17}$$

where μ_0 and μ_{∞} are the limiting rubbery and glassy moduli, f_0 is the frequency where the transition occurs, and α and β are two constants. However, at the low frequencies consid-

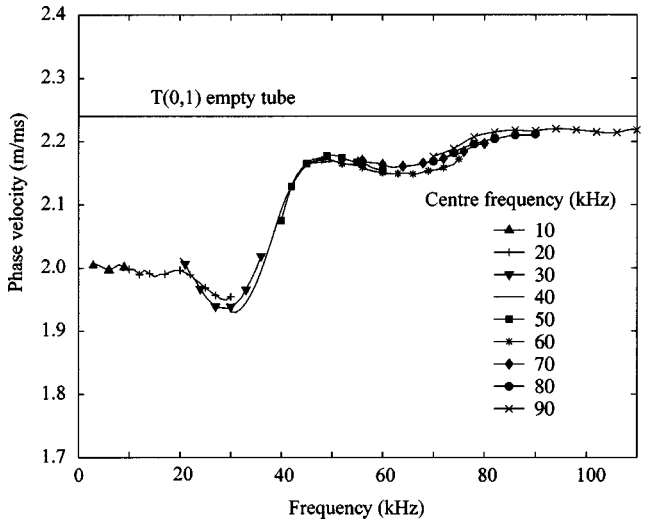


FIG. 7. Experimental phase velocity versus frequency for a 6.8 mm inner radius, 0.7 mm wall thickness copper tube filled with bitumen.

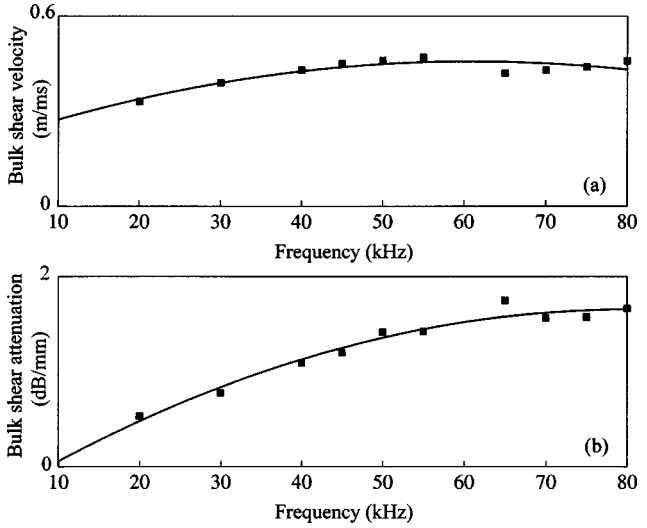


FIG. 9. Dispersion curves for bitumen TML 24515 45/60: (a) bulk shear velocity against frequency; (b) bulk shear attenuation against frequency.

TABLE I. Material bulk properties and geometry.

Material	Velocity (m/ms)	Attenuation (Np/wl)	Density (kg/m^3)	Thickness/radius (mm)
Copper tube	2.24	0.00	8900	0.70
Inner core	0.43	1.35	970	6.80

ered in this paper, due to the large value of f_0 (typically 2 MHz), the real part of $\tilde{\mu}$ is almost constant with frequency, while the imaginary part varies linearly with frequency since the ratio f/f_0 is very low. This suggests that a polynomial interpolation, at low frequencies, should be able to describe with an acceptable level of accuracy the material behavior, and so avoiding a time-consuming least squares fit to the measured data with Eq. (17).

VI. VALIDATION

In order to evaluate the accuracy of the method, the effect of errors in the phase velocity and guided wave attenuation measurements on the bulk properties evaluation has been studied numerically. For a system with the properties shown in Table I, it has been found that the best sensitivity is reached around the frequency of the first guided wave attenuation peak over a frequency range of 60% of the peak frequency (60% bandwidth). In particular, for a 10% guided wave attenuation error and a 1% phase velocity error, which are representative of the level of confidence of the measurements, the maximum bulk shear velocity and bulk attenuation errors are about 4% and 20%, respectively. However, the errors at the frequency of maximum guided wave attenuation are considerably lower (2% and 12% in the bulk velocity and bulk attenuation estimates, respectively).

The lack of data on the acoustic properties of viscoelastic materials makes the validation of the technique very difficult. One possibility would be the use of viscous liquids, such as glycerol whose viscosity data are available in literature. However, due to the low value of the viscosity, the shear impedance of glycerol is much lower than that of any metal. As a consequence, the amount of energy which can be transferred from the tube into the fluid core is very low. This leads to a low value of the guided wave attenuation which is difficult to measure unless a very long tube is used. On the other hand, high viscosity standards are extremely sensitive to temperature changes. For instance, Polybutene viscosity varies from 36.2 to 18.3 Pa s when the temperature ranges between 293 and 299 K. As a result, an extremely accurate temperature control would be required. Moreover, some concerns arise about the validity of the Newtonian model for these large values of the viscosity.

A partial validation of the technique comes from the fact that the dispersion curves obtained for bitumen are in reasonable agreement with the Kramers–Krönig relationship,[24] according to which the bulk shear velocity spectrum can be derived from the bulk attenuation spectrum

$$\frac{1}{c_0}-\frac{1}{c(\omega)}=\frac{2}{\pi}\int_{\omega_0}^{\omega}\frac{\alpha(\omega')}{\omega'^2}d\omega', \tag{18}$$

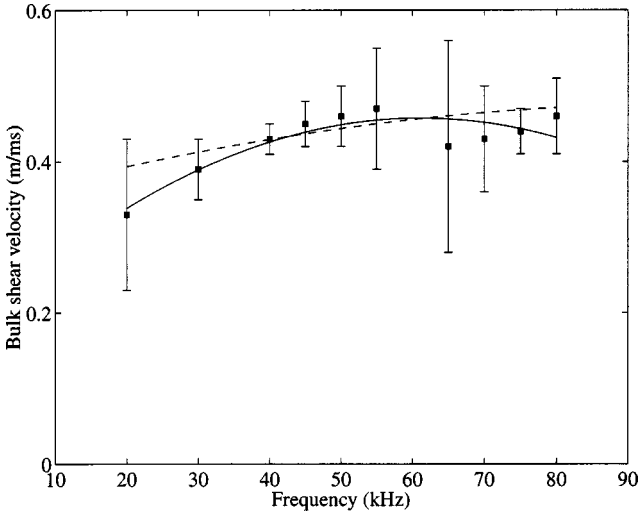


FIG. 10. Bulk velocity dispersion curve for bitumen TML 24515 45/60: (■) experiments; (—) experimental data interpolation; (---) Kramers–Krönig formula.

where c_0 is the velocity at the reference angular frequency ω_0 and α is the bulk attenuation in Nepers per unit length. Figure 10 shows the measured bulk shear velocity and the prediction from the Kramers–Krönig formula applied to the experimental bulk attenuation spectrum shown in Fig. 9(b). The error bars indicate the level of confidence of the measured bulk velocities; the estimated errors vary with frequency as discussed earlier. The reference frequency in the integral (18) was chosen to be equal to the frequency of maximum guided wave attenuation (40 kHz) since, at this frequency, the level of confidence of both bulk attenuation and bulk velocity is optimum. The measured bulk velocities (squares) fit well with the Kramers–Krönig formula prediction (dashed line) for frequencies ranging between 30 and 55 kHz. However, at higher frequency, the fitting becomes less accurate. This is due to the fact that, close to the frequency of the second guided wave attenuation minimum, the sensitivity drops as shown by the error bars in Fig. 10. The bulk velocity for a simple viscoelastic material should increase monotonically with frequency whereas Fig. 10 suggests that the velocity has a local maximum at around 60 kHz. This apparent maximum is probably due to the increased error in the estimates around the second guided wave attenuation minimum.

VII. MAXIMA AND MINIMA IN THE GUIDED WAVE ATTENUATION SPECTRUM

The occurrence of maxima and minima in the guided wave attenuation spectrum can be explained by considering the dispersion curves of a filled tube in which the inner core is perfectly elastic.

Figure 11(a) shows the dispersion curves for the filled tube described in Table I when the bulk attenuation of the core is neglected. For an empty tube, in the low frequency range, the only propagating mode is $T(0,1)$ [dashed line in Fig. 11(a)]. This mode is nondispersive and travels without being damped at a phase velocity which coincides with the shear velocity of the tube (see, for instance, Rose[25]). How-

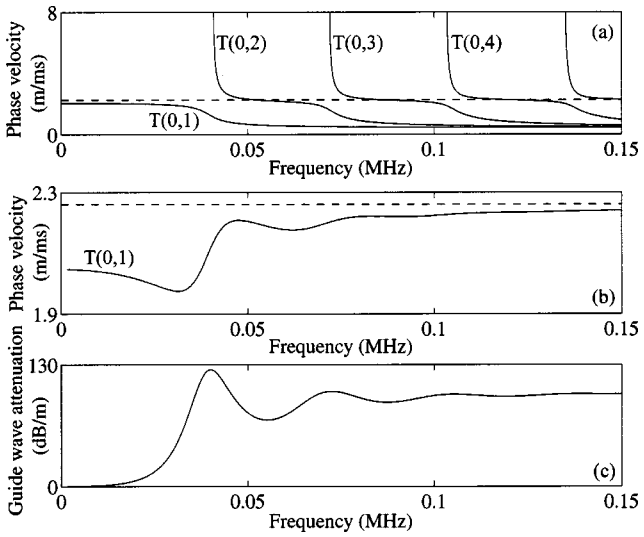


FIG. 11. Dispersion curves for the system described in Table I: (a) Phase velocity of the filled tube when the bulk attenuation is neglected (solid line) and phase velocity of the empty tube (dashed line); (b) as (a) but with bulk attenuation in the core material; (c) guided wave attenuation corresponding to (b).

ever, when the tube is filled $T(0,1)$ is replaced by an infinite number of new modes as can be deduced from Fig. 11(a). At the cutoff frequencies of these modes, due to the large impedance difference between the core and the tube, the modes correspond to standing waves in the core, while there is very little motion of the tube wall. However, as the frequency increases, the amount of energy which travels in the tube wall increases. At the frequencies where the modes intersect the $T(0,1)$ mode of the empty tube, most of the energy propagates in the tube rather than in the core. As the frequency increases further, the energy contained in the tube wall decreases, and for very high frequency the energy primarily travels in the core.

Figures 11(b) and (c) show the phase velocity and guided wave attenuation for a viscoelastic core with constant bulk shear velocity and attenuation per wavelength (see Table I) [note the velocity scale in Fig. 11(b) is finer than that in Fig. 11(a)]. In this case, as the frequency increases, the $T(0,1)$ mode of the filled tube approaches the $T(0,1)$ mode of the empty tube rather than tending to the $T(0,1)$ mode of the free core. All the other higher order modes of the filled tube are strongly damped,[26] and for this reason of no practical interest. It is interesting to note that peaks in the guided wave attenuation occur around the cutoff frequencies of the elastic filled tube, Figs. 11(a) and (c). It can be shown that the guided wave attenuation is proportional to the strain energy contained in the viscoelastic core per unit axial power flow.[26] Since at the cutoff frequencies of the elastic case the tube vibrations are negligible, it follows that at these frequencies strain energy maxima in the core have to occur. As a consequence, guided wave attenuation maxima occur at the cutoff frequencies of the equivalent elastic filled tube. Similarly, guided wave attenuation minima are due to the fact that, at these frequencies, the energy flows mainly in the tube wall. As a result, the strain energy content of the core per unit power flow is a minimum.

VIII. CONCLUSIONS

A novel technique for the characterization of highly attenuative viscoelastic materials has been presented. It has been shown that by measuring the dispersion curves of a tube filled with the unknown material the shear acoustic properties can be obtained. The method is based on the hypothesis of linear viscoelasticity and no assumptions are made about the frequency dependence of the acoustic properties as a monochromatic approach is followed.

The technique is attractive since many of the limitations common to traditional methods including beam spreading, sample manufacturing, transducer coupling, etc., are overcome. On the other hand, the need for the material to be molded into the tube represents a limitation in the case of solid materials, since melting the material would dramatically change the properties. By contrast, the method is particularly suitable for materials that will flow as the tube can then easily be filled.

[1] D. N. Alleyne, B. Pavlakovic, M. J. S. Lowe, and P. Cawley, "Rapid, long range inspection of chemical plant pipework using guided waves," Insight **43**, 93–96 (2001); **43**, 101 (2001).

[2] D. K. Mak, "Comparison of various methods for the measurement of reflection coefficient and ultrasonic attenuation," Br. J. Non-Destr. Test. **33**, 441–449 (1991).

[3] D. K. Mak, "Ultrasonic phase velocity measurement incorporating couplant correction," Br. J. Non-Destr. Test. **35**, 443–449 (1993).

[4] R. E. Challis, G. P. Wilkinson, and F. R. J., "Errors and uncertainties in the ultrasonic pulse-echo reflectometry method for measuring acoustic impedance," Meas. Sci. Technol. **9**, 692–700 (1998).

[5] M. Castaings, B. Hosten, and T. Kundu, "Inversion of ultrasonic, plane-wave transmission data in composite plates infer viscoelastic material properties," NDT & E International **33**, 377–392 (2000).

[6] A. Di Meglio and L. S. Wang, "A variational method for identification of viscoelastic parameters from experimental data," J. Acoust. Soc. Am. **6**, 2476–2753 (2000).

[7] T. Pialucha, C. C. H. Guyott, and P. Cawley, "Amplitude spectrum method for the measurement of phase velocity," Ultrasonics **27**, 270–279 (1989).

[8] N. Guo, M. K. Lim, and T. Pialucha, "Measurement of the attenuation using a normalized amplitude spectrum," J. Nondestruct. Eval. **14**, 9–19 (1995).

[9] W. Sachse and Y. H. Pao, "On the determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **64**, 1645–1651 (1987).

[10] P. Kielczynski and R. Plowiec, "Determination of the shear impedance of viscoelastic liquids using Love and Bleustein-Gulyaev surface waves," J. Acoust. Soc. Am. **86**, 818–827 (1989).

[11] J. O. Kim, Y. Wang, and H. H. Bau, "The effect of an adjacent viscous fluid on the transmission of torsional stress waves in a submerged waveguide," J. Acoust. Soc. Am. **89**, 1414–1422 (1991).

[12] T. Vogt, M. Lowe, and P. Cawley, "Ultrasonic waveguide techniques for the measurement of material properties," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, New York, 2002), Vol. 21, pp. 1742–1749.

[13] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. 2.

[14] Y. M. Haddad, *Viscoelasticity of Engineering Materials* (Chapman and Hall, New York, 1995).

[15] R. M. Christensen, *Theory of Viscoelasticity: An Introduction* (Academic, New York, 1971).

[16] G. Caviglia and A. Morro, "A new approach to reflection-transmission between viscoelastic half-spaces," J. Acoust. Soc. Am. **106**, 1666–1672 (1999).

[17] M. J. S. Lowe, "Matrix techniques for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).

[18] B. N. Pavlakovic, M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "DISPERSE: A general purpose program for creating dispersion curves," in

Review of Progress in Quantitative NDE, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 185–192.
[19] J. Vollman and J. Dual, ''High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium. I. Theory and numerical results,'' J. Acoust. Soc. Am. **102**, 896–908 (1997).
[20] J. Vollman, R. Breu, and J. Dual, ''High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium. II. Experimental results versus theory,'' J. Acoust. Soc. Am. **102**, 909–920 (1997).
[21] J. M. Carcione and G. Seriani, ''Torsional waves in lossy cylinders,'' J. Acoust. Soc. Am. **103**, 760–766 (1998).
[22] R. B. Randall, *Frequency Analysis* (Bruel&Kjaer, Denmark, 1995).
[23] B. Hartmann, G. F. Lee, and J. D. Lee, ''Loss factors height and with limits for polymer relaxations,'' J. Acoust. Soc. Am. **95**, 226–233 (1994).
[24] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers-Kronig relationship between ultrasonic attenuation and phase velocity,'' J. Appl. Mech. **69**, 696–701 (1981).
[25] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, 1999).
[26] F. Simonetti and P. Cawley, ''On the nature of shear horizontal wave propagation in elastic plates coated with viscoelastic materials,'' Proc. R. Soc. London (submitted).

Theory of the Greenspan viscometer

Keith A. Gillis,[a] James B. Mehl,[b] and Michael R. Moldover[c]
Process Measurements Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8360

(Received 14 November 2002; accepted for publication 7 April 2003)

We present a detailed acoustic model of the Greenspan acoustic viscometer, a practical instrument for accurately measuring the viscosity η of gases. As conceived by Greenspan, the viscometer is a Helmholtz resonator composed of two chambers coupled by a duct of radius r_d. In the lowest order, $\eta = \pi f \rho (r_d/Q)^2$, where f and Q are the frequency and quality factor of the isolated Greenspan mode, and ρ is the gas density. In this level of approximation, the viscosity can be determined by measuring the duct radius and frequency response of the resonator. In the full acoustic model of the resonator, the duct is represented by a *T*-equivalent circuit, the chambers as lumped impedances, and the effects of the diverging fields at the duct ends by lumped end impedances with inertial and resistive components. The model accounts for contributions to $1/Q$ from thermal dissipation (primarily localized in the chambers) and from a capillary used for filling and evacuating the resonator. A robust, prototype instrument is being used for measuring the viscosity of reactive gases used in semiconductor processing. For well-characterized surrogate gases, the prototype viscometer generated values of η that were within $\pm 0.8\%$ of published reference values throughout the pressure range 0.2–3.2 MPa. Remarkably, we achieved this level of agreement by only slight adjustment of the numerically calculated inertial and resistive end effect parameters to improve the agreement with helium reference values. No other parameters were adjusted. [DOI: 10.1121/1.1577550]

PACS numbers: 43.35.Ud, 43.20.Ye [RR]

I. INTRODUCTION

As conceived by Greenspan,[1] the acoustic viscometer is a Helmholtz resonator composed of two gas-filled chambers connected by a duct of radius r_d. (See Fig. 1.) By design, the radius of the duct is much smaller than its length L_d and smaller than the dimensions of the chambers (radius r_c, length L_c). Such a resonator has a low frequency (Greenspan) mode in which the gas oscillates between the chambers through the duct. It is easy to measure the frequency response of the Greenspan mode very accurately because it is well below the frequencies of all other acoustic modes of the enclosed gas and the elastic modes of the resonator body. The damping of the Greenspan mode is determined primarily by the viscosity of the gas η; in the lowest order, $\eta = \pi f \rho (r_d/Q)^2$, where f and Q are the frequency and quality factor of the mode and ρ is the gas density.

In 1953, Greenspan and Wimenitz[1] constructed five prototype viscometers, and used these to measure the viscosity of air. Their results differed from accepted values by -1% to $+38\%$. They judged the results unsatisfactory and did not pursue development of the method.

Previously,[2] we reported measurements with two viscometers similar to those used by Greenspan and Wimenitz, and gave a brief description of the theoretical model used for analyzing the data. Measurements of the viscosities of helium, argon, and propane with one viscometer systematically differed from reference values by about 1%; measurements with the other viscometer differed from reference values by about 3%. These viscometers were evacuated and filled through a capillary connected to one of the chambers. Practical use of a viscometer in a thermally controlled sample cell requires a capillary whose length exceeds the dimensions of the viscometer. In this paper we show that the effects of long capillaries of this type on the viscometer response limits the accuracy of the instrument.

These considerations led to an improved design[3–5] with the capillary connected to the center of the resonator duct, where the coupling of the acoustic fields to the fill capillary are very small.

In this paper, we develop complete models of both types of viscometer. The low-frequency limit of Kirchhoff's[6] theory[7] is used to model acoustic flow in the main viscometer duct and the fill capillary. Duct end effects are described using the numerical results of Mehl.[8,9]

Our objective is an accurate frequency-response function so that the Greenspan viscometer can be used to determine accurately the viscosity of gases. Ultimately, we are concerned with the viscosity of the highly reactive gases used in semiconductor processing. The data are needed to calibrate mass-flow controllers and to model processes such as chemical vapor deposition. Also, the viscosity of helium–argon and helium–xenon mixtures is needed to optimize thermoacoustic machinery.

The theoretical model was tested by analyzing results from a prototype Greenspan viscometer using five well-studied surrogate gases over the pressure range 0.2–3.2 MPa.

[a]Electronic mail: keith.gillis@nist.gov
[b]Electronic mail: jmehl@rockisland.com. Also at: 36 Zunuqua Trail, P.O. Box 307, Orcas, WA 98280-0307.
[c]Electronic mail: michael.moldover@nist.gov

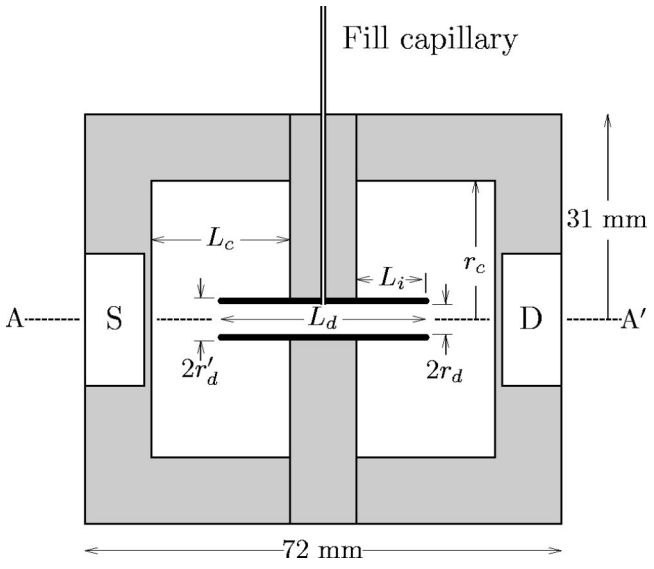


FIG. 1. Cross section of a Greenspan viscometer. The dotted line AA′ indicates the axis of circular symmetry for all parts except the fill capillary. Two cylindrical chambers fitted with source (*S*) and detector (*D*) transducers are coupled by a concentric cylindrical duct. The dimensions in millimeters are $L_d=31.17$, $r_d=2.316$, $r'_d=3.21$, $r_c=21.02$, $L_c=21.04$, and $L_i=10.5$. The fill capillary has an inner radius $r_f=0.10$ mm (not to scale) and a length $L_f=80$ mm. As shown, it is attached at the center of the duct. The viscometers described in Ref. 2 have a similar shape, except that that $L_i=0$ and the fill capillary is attached at the wall of one of the chambers.

The resulting values of η were within $\pm 0.8\%$ of published reference values.

Some of the notation used in this paper is summarized below. Other quantities are defined in the text,

$\rho=$ mass density,

$p=$ acoustic pressure,

$c=$ speed of sound,

$C_p(C_v)=$ heat capacity per unit mass at constant pressure (volume),

$\gamma=C_p/C_v$,

$\lambda=$ thermal conductivity,

$D_v=\eta/\rho=$ viscous diffusivity,

$D_t=$ thermal diffusivity $=\lambda/(\rho C_p)$,

$\delta_v=\sqrt{2D_v/\omega}=$ viscous penetration length,

$\delta_t=\sqrt{2D_t/\omega}=$ thermal penetration length.

All ''small'' acoustic quantities have an implicit time dependence proportional to $e^{i\omega t}$.

A. Resonator description

Figure 1 shows the Greenspan viscometer described by Wilhelm *et al.*[3–5] Two right-circular cylindrical chambers coupled by a concentric cylindrical duct form a resonator whose low-frequency mode consists of oscillatory gas flow between the chambers (assumed to be of equal volume). The ''mass element'' of the resonator is the cylindrical mass of gas in the duct and near the duct ends. It has a nearly uniform flow velocity and effective mass $A_d(L_d+2\delta_i)\rho$, where ρ is the gas density, $A_d=\pi r_d^2$ is the cross-sectional area of the duct, L_d is the duct length, and the length $\delta_i\approx 0.655\cdot r_d$ is an inertial end correction which accounts for diverging flow at the duct ends. Flow into a chamber increases the pressure in that chamber and provides a restoring force; the effective force constant is $\rho c^2 A_d^2/V_c$ per chamber, where c is the speed of sound and V_c is the volume of a single chamber. A zero-order estimate of the frequency $\omega_0=ck_0$ is

$$k_0^2=\frac{2A_d}{V_c(L_d+2\delta_i)}. \tag{1}$$

(If the chambers have unequal volumes V_1 and V_2, $2/V_c$ should be replaced by $1/V_1+1/V_2$.) For the resonator used in this work, Eq. (1) has a numerical value $k_0\approx 5.8\,\mathrm{m}^{-1}$. The corresponding frequency is a few hundred Hz for typical gases, well below the frequencies of elastic vibrations of the resonator body and higher acoustic modes of the gas in the resonator. Thus, the Helmholtz mode is isolated and nondegenerate, which is desirable if η is to be determined with the highest accuracy from measurements of the mode's resonance response.

We did not use the viscometer below 100 kPa. As the pressure is decreased, the signal-to-noise ratio decreases because of decreasing Q and increasing impedance mismatches with the transducers. At 300 K and 100 kPa, the ratio of the mean free path to the viscous penetration length δ_v was about 2×10^{-4} for helium and propane, and about 5×10^{-4} for argon. Thus inertial and thermal accommodation effects in the boundary conditions can be ignored.

B. Principle of the measurement

In this section, we describe the principle of the measurement by considering the viscous and other contributions to the inverse quality factor $(1/Q)$ of the resonator.

Acoustic energy dissipation in the gas is a result of both surface phenomena (energy transport near the resonator wall) and volume phenomena (energy transport far from the wall). At low frequencies and at pressures in the range 0.1 MPa to 3 MPa, the main sources of dissipation for nonrelaxing gases are viscous and thermal boundary layer phenomena, which can be estimated using the formulas derived by Morse and Ingard for flat surfaces.[10] The rate of kinetic energy loss within the viscous boundary layers in the duct is

$$\dot{E}_v=-\tfrac{1}{2}\rho\omega\delta_v u_d^2\cdot 2\pi r_d(L_d+2\epsilon_r r_d), \tag{2}$$

where u_d is the rms value of the gas velocity in the duct and $\epsilon_r r_d$ is an effective length associated with viscous dissipation at the duct ends. The rate of potential energy loss within the thermal boundary layer in the chambers is

$$\dot{E}_t=-\frac{(\gamma-1)\omega\delta_t}{2\rho c^2}p_{ch}^2\cdot 2S_c, \tag{3}$$

were p_{ch} is the rms pressure amplitude in the chambers and S_c is the total surface area within one chamber. The total acoustic energy in the resonator can be written either as kinetic energy or potential energy, the former approximately

localized in the duct, the latter in the chambers:

$$E \approx \rho u_d^2 \cdot A_d (L_d + 2\delta_i) \approx \frac{p_{ch}^2}{\rho c^2} \cdot 2V_c . \tag{4}$$

Acoustic dissipation within the resonator volume is important for gases at very low density, gases near the liquid-vapor critical point, and relaxing gases under a wide range of conditions. The fractional rate of energy loss due to these processes is

$$\frac{\dot{E}_{\text{volume}}}{E} = -\frac{\omega^2}{c^2}\left[\frac{4}{3} D_v + (\gamma - 1) D_t + \frac{\eta_b}{\rho}\right]. \tag{5}$$

The first two terms in Eq. (5) are the classical absorption terms. The third term describes absorption due to relaxation processes and is characterized by the bulk viscosity η_b. Relaxation processes govern the energy exchange between translational degrees of freedom (the acoustic mode) and internal modes of the fluid, such as molecular vibrations, metastable intermolecular bound states, or critical fluctuations. If the process can be characterized by a single relaxation time τ_{relax}, then the bulk viscosity is approximately

$$\eta_b = (\gamma - 1)\rho c^2 \frac{C_{\text{relax}}}{C_p} \frac{\tau_{\text{relax}}}{1 + (\omega \tau_{\text{relax}})^2}. \tag{6}$$

The inverse resonance quality factor $1/Q = -\dot{E}/(\omega E)$ is the sum of three terms:[11]

$$\frac{1}{Q} = \frac{1}{Q_v} + \frac{1}{Q_t} + \frac{1}{Q_{\text{volume}}}, \tag{7}$$

$$\frac{1}{Q_v} \equiv \frac{\delta_v}{r_d} \cdot \frac{L_d + 2\epsilon_r r_d}{L_d + 2\delta_i}, \tag{8}$$

$$\frac{1}{Q_t} \equiv \frac{(\gamma - 1)\delta_t S_c}{2V_c}, \tag{9}$$

$$\frac{1}{Q_{\text{volume}}} \equiv \frac{\omega^2}{2c^2}\left[\frac{4}{3}\delta_v^2 + (\gamma - 1)\delta_t^2 + \frac{2\eta_b}{\rho\omega}\right]. \tag{10}$$

If end effects are neglected, $1/Q_v$ is simply δ_v / r_d. For the resonator in Fig. 1, the numerical value of ϵ_r is approximately 0.95 under typical experimental conditions, so the dissipative end effect acts to increase the viscous dissipation by an amount $2\epsilon_r r_d / L_d \approx 0.14$. The inertial end effect acts to increase the kinetic energy and hence decreases $1/Q_v$ by an amount $2\delta_i / L_d \approx 0.097$. Together, the inertial and dissipative end effects in the resonator contribute a net increase in $1/Q_v$ by a factor of 1.04.

Equations (7) and (8) show that when thermal boundary layer and bulk absorption can be neglected, the viscosity can be determined as

$$\eta = \frac{1}{2}\omega\rho\delta_v^2 = \frac{\pi f \rho r_d^2}{Q^2}\left(\frac{L_d + 2\delta_i}{L_d + 2\epsilon_r r_d}\right)^2. \tag{11}$$

The only requirements to measure f and Q are transducers that (1) have a smooth frequency response (no sharp peaks) in the vicinity of the Helmholtz mode and (2) are stable during the few minutes required to measure the frequency response of the Helmholtz mode. The quality factor should be small enough so that $1/Q$ can be measured accurately but not so small that signal/noise is seriously degraded. In practice, these considerations lead to designs that work in the range $20 < Q < 100$.

For chambers with equal radii and lengths ($r_c = L_c$), the ratio V_c/S_c in Eq. (9) is $r_c/4$. Because the penetration lengths δ_v and δ_t in gases are typically of the same magnitude, the desired dominance of the viscous term in Eq. (7) requires $r_d \ll \frac{1}{2} r_c/(\gamma - 1)$. An attempt to satisfy this criterion drives the design toward cumbersome large chambers, low frequencies, and low Qs, where measurements become difficult. Thus, for the practical, compact acoustic viscometer shown in Fig. 1, the thermal diffusivity contribution to Eq. (11) must be included. At 300 K and 0.2 MPa the thermal term is 36% of the viscous term for argon and 7% for propane. Also, the advantages of a reasonably compact resonator require a duct of moderate length. For such a duct, the parameter ϵ_r must be either calculated or determined from measurements.

II. ACOUSTIC MODEL OF THE RESONATOR

A. Acoustic propagation in a circular duct

Acoustic waves in gas-filled ducts are governed by the equations first proposed by Kirchhoff,[6] whose classic paper includes a description of the effects of the coupled acoustic, thermal, and vorticity waves in ducts of circular cross section. The low-frequency limit of his solutions, in a form appropriate for this paper, is generally attributed to Crandall.[7] We assume an implicit time dependence of all "small" acoustic quantities proportional to $e^{i\omega t}$ in the following. The acoustic pressure and volume velocity U are related by the transmission line equations

$$\frac{\partial p}{\partial z} = -\mathcal{Z} U, \quad \frac{\partial U}{\partial z} = -\mathcal{Y} p, \tag{12}$$

where $\mathcal{Z}$ is the series impedance and $\mathcal{Y}$ is the parallel admittance. For a circular duct in the low-frequency limit, these quantities are

$$\begin{aligned} \mathcal{Z} &= \frac{i\omega\rho}{A_d}\frac{1}{1 - \mathbf{F}_v}, \\ \mathcal{Y} &= \frac{i\omega A_d}{\rho c^2}[1 + (\gamma - 1)\mathbf{F}_t], \end{aligned} \tag{13}$$

where $\mathbf{F}_v$ and $\mathbf{F}_t$ are defined by

$$\mathbf{F}_x = \frac{2J_1(\kappa_x r_d)}{\kappa_x r_d J_0(\kappa_x r_d)} \tag{14}$$

in terms of the Bessel functions $J_m(\zeta)$ and the parameters $\kappa_v = (1 - i)/\delta_v$ and $\kappa_t = (1 - i)/\delta_t$.

Elimination of U from Eq. (12) yields

$$\frac{\partial^2 p}{\partial z^2} = \mathcal{Z}\mathcal{Y} p = \mathbf{\Gamma}^2 p, \tag{15}$$

where $\mathbf{\Gamma} = \sqrt{\mathcal{Z}\mathcal{Y}}$ is the propagation parameter for waves along the duct. Waves propagating in the $\pm z$ directions have

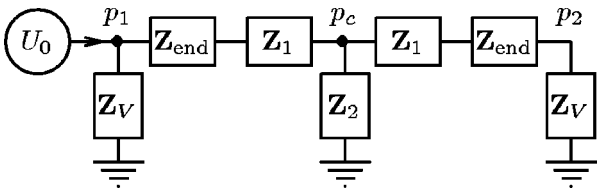


FIG. 2. Simplified equivalent circuit of the Greenspan viscometer.

a z-dependence proportional to $e^{\pm\Gamma z}$. It follows that

$$U=\pm\frac{\Gamma}{\mathcal{Z}}p=\pm\frac{\mathcal{Y}}{\mathcal{Z}}p=\pm\frac{p}{\mathbf{Z}_0}, \tag{16}$$

where $\mathbf{Z}_0=\sqrt{\mathcal{Z}/\mathcal{Y}}$ is the characteristic impedance. In a lumped-component model, finite-length transmission lines can be replaced by T-equivalent circuits.[12] Figure 2 shows the equivalent circuit of the viscometer duct terminated by the impedances associated with end effects and chambers (as defined in the next section). The impedances used to represent the duct are

$$\mathbf{Z}_1=\mathbf{Z}_0\tanh(\mathbf{\Gamma}L_d/2)=\frac{i\omega\rho L_d}{2A_d(1-\mathbf{F}_v)}\frac{\tanh(\frac{1}{2}\mathbf{\Gamma}L_d)}{\frac{1}{2}\mathbf{\Gamma}L_d}, \tag{17}$$

$$\mathbf{Z}_2=\frac{\mathbf{Z}_0}{\sinh(\mathbf{\Gamma}L_d)}=\frac{\rho c^2}{i\omega A_d L_d[1+(\gamma-1)\mathbf{F}_t]}\frac{\mathbf{\Gamma}L_d}{\sinh(\mathbf{\Gamma}L_d)}. \tag{18}$$

In each case, the second form involves the ratio of a hyperbolic function to its argument, ratios which approach unity at zero frequency.

B. Lumped-component model

Lumped component acoustic models have been used for analyzing experimental results.[2–5] The use of such models is supported by numerical calculations,[8] which show that the pressure field within each chamber is spatially uniform except for entrance effects at the duct ends. The entrance effects are confined to a region whose dimensions are of order r_d. The effects of the chambers can be represented as a volume impedance in series with an end-effects term. Within the chambers the gas compressibility is isothermal at the walls and adiabatic far from the walls. An expression for the chamber impedance obtained by calculating the average compressibility within the chamber is[13]

$$\mathbf{Z}_V=\frac{\rho c^2}{i\omega[V_c+\frac{1}{2}(1-i)(\gamma-1)S_c\delta_t/(1+\chi_t)]}=\frac{\rho c^2}{i\omega\mathbf{V}_c}, \tag{19}$$

where $\chi_t=\sqrt{\rho_g C_{pg}\lambda_g/(\rho_s C_{ps}\lambda_s}$ accounts for thermal waves in the solid walls of the chamber. The subscripts on the densities, specific heats, and thermal conductivities in this expression refer to the gas and solid. For the stainless steel chambers used in this work, the largest value of χ_t was 0.008 for helium 3.2 MPa. A complex volume $\mathbf{V}_c$ has been introduced in the second form for brevity. The ratio $S_c\delta_t/V_c$ is the approximate fraction of the chamber's volume in which the compressibility differs from the adiabatic value and where dissipation occurs. Equation (19) applies to a rigid chamber, and is subject to modification due to the transducer impedance and elastic deformation of the chamber body. The frequency response of the measured shell was measured and numerous modes were found above 12 kHz. Analysis of the mechanical parts show that the effects of the transducer impedance are negligible, but that the compliance of the screws holding the viscometer together introduce a fractional correction to Eq. (19) on the order of $10^{-4}\,p_0/\text{MPa}$, where p_0 is the sample pressure. This may be the source of our observations of a small, spurious pressure dependence of the speed of sound.

The localized impedance at the duct ends can be written

$$\mathbf{Z}_{\text{end}}=\frac{\rho\omega}{A_d}(i\,\delta_I+\delta_R), \tag{20}$$

where δ_I and δ_R are lengths which have been evaluated numerically.[8] To first order in δ_v the resistive length is

$$\delta_R=\epsilon_r\delta_v\,. \tag{21}$$

When the duct ends are square, the techniques of Ref. 8 predict a constant resistive parameter $\epsilon_{r0}\approx 0.987$. The boundary-layer approximation used in deriving this number breaks down when the tangential velocity varies rapidly on a length scale of order δ_v, as it does at the sharp corners of the duct ends where the acoustic velocity field is weakly divergent. Cutting off the singular integrals a distance of order δ_v from the sharp corners yields a fractional decrease in ϵ_r of order $(\delta_v/r_d)^{1/3}$. Recent numerical calculations[9] of the acoustic and vorticity fields near the duct end show that ϵ_r is slightly less than the constant term and varies weakly with δ_v/r_d according to

$$\epsilon_r\approx\epsilon_{r0}-0.348(\delta_v/r_d)^{1/3}+1.15(\delta_v/r_d). \tag{22}$$

The constant term in this expression is sensitive to the shape of the duct ends. Rounding the sharp edges decreases ϵ_r, and a slight burr can increase it. We increased ϵ_{r0} by 4% to 1.03 to improve the agreement of the experimental values of the viscosity of helium with the accurately known theoretical values.[14] With this value of ϵ_{r0}, Eq. (22) predicts values of ϵ_{r0} varying from 0.97 to 0.96 for the range $0.008<\delta_v/r_d<0.045$ that corresponds to the data that we present in Appendix A. Those data also support the weak dependence of ϵ_r of Eq. (22). In future work, we recommend rounding the corners of the duct ends to reduce the sensitivity of ϵ_r to δ_v/r_d.

The inertial length δ_I in Eq. (20) is, to first order in δ_v,

$$\delta_I=\delta_i+\epsilon_i\delta_v\,. \tag{23}$$

As noted previously, $\delta_i\approx 0.655\cdot r_d$ for the experimental resonator discussed here; δ_I differs from this by a viscous boundary layer correction term and hence provides a more accurate representation of inertial end effects. The argument in Ref. 8 that $\epsilon_i\approx\epsilon_r$ failed to take into account the directionality of the linear momentum, an effect that lowers the portion of ϵ_i associated with diverging flow outside the duct end by a small amount. The value $\epsilon_i\approx 0.86$ obtained in recent numerical work[9] is slightly smaller than ϵ_{r0}. Improved experimental agreement was obtained with a modest increase to $\epsilon_i\approx 0.96$.

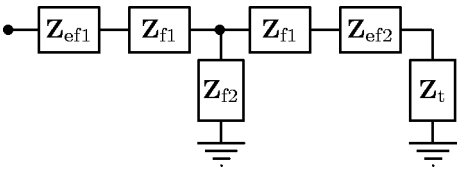


FIG. 3. Equivalent circuit of a fill capillary, treated as a circular duct terminated by an impedance $\mathbf{Z}_t$.

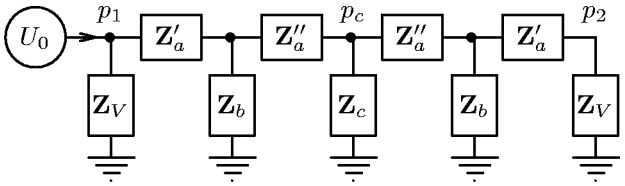


FIG. 4. Modified equivalent circuit with centered fill capillary represented by impedance $\mathbf{Z}_c$.

It is convenient to first consider a resonator without a fill capillary, as shown in Fig. 2. Equation (18) shows that $\mathbf{Z}_2$ is proportional to the inverse duct volume, so it is much larger than $\mathbf{Z}_V$, which is proportional to the inverse chamber volume. Accordingly, forced acoustic flow will be mainly between the chambers with small perturbations due to the $\mathbf{Z}_2$ term.

The Greenspan viscometer has a low-frequency mode of odd symmetry which is completely independent of $\mathbf{Z}_2$. For this mode, the acoustic pressures in Fig. 2 satisfy $p_2=-p_1$ and $p_c=0$. This will occur if the sum of the impedances in either leg of the "*T*" vanishes

$$\mathbf{Z}_1+\mathbf{Z}_{\text{end}}+\mathbf{Z}_V=0. \tag{24}$$

This simplifies to

$$\frac{\omega^2}{c^2}=\frac{2A_d(1-\mathbf{F}_v)/\mathbf{V}_c}{L_d\tanh(\frac{1}{2}\boldsymbol{\Gamma}L_d)/(\frac{1}{2}\boldsymbol{\Gamma}L_d)+2(\delta_I-i\delta_R)(1-\mathbf{F}_v)}. \tag{25}$$

In the absence of dissipative terms this agrees exactly with Eq. (1), except for the hyperbolic-tangent term in the denominator. The magnitude of this term can be estimated by using $\boldsymbol{\Gamma}\approx i\omega/c\approx ik_0$, so that $(\boldsymbol{\Gamma}L_d)^2\approx-2A_dL_d/V_c$, and

$$\frac{\tanh(\frac{1}{2}\boldsymbol{\Gamma}L_d)}{(\frac{1}{2}\boldsymbol{\Gamma}L_d)}\approx 1+\frac{2A_dL_d}{3V_c}. \tag{26}$$

This differs from unity by 2/3 of the ratio of the duct volume to a chamber volume, or 1.2% for the experimental resonator. Terms like this are not negligible in the full model, but can be neglected in Eq. (25) for present purposes. For small dissipation, neglecting terms of second order in δ_v and δ_t, approximating $\mathbf{F}_v$ by $(1-i)\delta_v/r_d$, and writing $\omega^2\approx\omega_1^2(1+i/Q)$ where ω_1 is real, Eq. (25) simplifies further to

$$\frac{\omega_1^2}{c^2}\left(1+\frac{i}{Q}\right)\approx k_0^2\left(1+\frac{i}{Q_v}+\frac{i-1}{Q_t}-\frac{L_d+2\epsilon_i r_d}{L_d+2\delta_i}\frac{\delta_v}{r_d}\right). \tag{27}$$

The imaginary part of this equation is identical with the boundary-layer contributions to $1/Q$ in Eq. (7); the real parts give an improved estimate of the resonance frequency.

When there is an acoustic source with volume velocity U_0 in chamber 1, the steady-state pressure in chamber 2 calculated from the equivalent circuit is

$$p_2=\frac{\mathbf{Z}_V^2\mathbf{Z}_2U_0}{(\mathbf{Z}_1+\mathbf{Z}_{\text{end}}+\mathbf{Z}_V)(\mathbf{Z}_1+\mathbf{Z}_{\text{end}}+\mathbf{Z}_V+2\mathbf{Z}_2)}, \tag{28}$$

which has the expected resonance term in the denominator. (The second factor in the denominator vanishes for the lowest plane-wave mode in the duct, with $kL_d\approx\pi$.)

C. Input impedance of fill capillary

The fill capillary is treated as a circular duct terminated by an impedance $\mathbf{Z}_t$ (Fig. 3). It has an input impedance

$$\mathbf{Z}_c=\mathbf{Z}_{ef1}+\mathbf{Z}_{f1}+\frac{\mathbf{Z}_{f2}(\mathbf{Z}_{f1}+\mathbf{Z}_{ef2})}{\mathbf{Z}_{f2}+\mathbf{Z}_{ef2}+\mathbf{Z}_{f1}} \tag{29}$$

when the termination is open $(\mathbf{Z}_t\approx 0)$ and

$$\mathbf{Z}_c=\mathbf{Z}_{ef1}+\mathbf{Z}_{f1}+\mathbf{Z}_{f2} \tag{30}$$

when the termination is closed. In these expressions the impedances $\mathbf{Z}_{f1}$ and $\mathbf{Z}_{f2}$ are defined by Eqs. (17) and (18) with parameters appropriate for the fill capillary, i.e., $\boldsymbol{\Gamma}\rightarrow\boldsymbol{\Gamma}_f$, $L_d\rightarrow L_f$, and appropriate end corrections in Eq. (20) are used for $\mathbf{Z}_{ef1}$ and $\mathbf{Z}_{ef2}$.

D. Resonator response function

Early versions of the viscometer[2] had a fill capillary in the source chamber. In Appendix B, we derive the equation used for analyzing data taken with those viscometers, and show that it is difficult to minimize the effects of the fill capillary on the performance of the viscometer. These considerations led to relocating the fill capillary to the center of the duct in the viscometer shown in Fig. 1. The acoustic pressure has a node at this point for a symmetric resonator. This minimizes the coupling to the fill capillary, even for resonators with small asymmetries.

The modified viscometer is represented by the equivalent circuit in Fig. 4. Each half of the original duct is now represented by a *T* section. The impedance of the capillary duct, including end effects, is represented by $\mathbf{Z}_c$. The impedance $\mathbf{Z}'_a$ is the sum of the series term for a duct of length $L_d/2$ and the end impedance at the chamber

$$\mathbf{Z}'_a=\mathbf{Z}_0\tanh(\boldsymbol{\Gamma}L_d/4)+\mathbf{Z}_{\text{end}}. \tag{31}$$

Similarly, the term

$$\mathbf{Z}''_a=\mathbf{Z}_0\tanh(\boldsymbol{\Gamma}L_d/4)+\mathbf{Z}'_{\text{end}} \tag{32}$$

includes a different end impedance, as appropriate for a junction of the fill capillary with the main duct. The other term in the *T* sections is

$$\mathbf{Z}_b=\frac{\mathbf{Z}_0}{\sinh(\boldsymbol{\Gamma}L_d/2)}. \tag{33}$$

The circuit will have a low-frequency mode with $p_c=0$ and $p_1=-p_2$ if

$$\mathbf{Z}_V+\mathbf{Z}'_a+\frac{\mathbf{Z}_b\mathbf{Z}''_a}{\mathbf{Z}_b+\mathbf{Z}''_a}=0. \tag{34}$$

After some manipulation this reduces to

$$\mathbf{Z}_1+\mathbf{Z}_{\rm end}+\mathbf{Z}_V+\frac{\mathbf{Z}'_{\rm end}}{[\cosh(\mathbf{\Gamma}L_d/2)+\mathbf{Z}'_{\rm end}/\mathbf{Z}_b]\cosh(\mathbf{\Gamma}L_d/2)}=0, \tag{35}$$

which is identical to Eq. (24) when the impedance $\mathbf{Z}'_{\rm end}=0$. This impedance has not been calculated; however it is proportional to a length on the order of the radius of the fill capillary, and thus should be negligible. It is kept in the following equations for completeness, but was set to zero when analyzing the data presented in this paper.

The output pressure p_2, obtained by solving the system of linear equations represented by Fig. 4, is

$$p_2=U_0\frac{\mathbf{Z}_b^2\mathbf{Z}_c\mathbf{Z}_V^2}{D_1D_2}, \tag{36}$$

where

$$D_1=(\mathbf{Z}_b+\mathbf{Z}''_a)(\mathbf{Z}_V+\mathbf{Z}'_a)+\mathbf{Z}_b\mathbf{Z}''_a,$$

$$D_2=\mathbf{Z}_b\mathbf{Z}_V+2\mathbf{Z}_c(\mathbf{Z}_b+\mathbf{Z}_V)+\mathbf{Z}'_a(\mathbf{Z}_b+2\mathbf{Z}_c)+\mathbf{Z}''_a(\mathbf{Z}_b+\mathbf{Z}_V+\mathbf{Z}'_a).$$

Note that the vanishing of the denominator factor D_1 is equivalent to the resonance expression (34). Equation (36) is the theoretical resonance response function used in data analysis.

III. TESTS OF THE THEORY

The theory presented above was tested with data acquired using the resonator sketched in Fig. 1. Measurements were taken on five well-studied gases: helium, argon, nitrogen, propane, and sulfur hexafluoride. For these gases, the speed of sound, the thermal conductivity, and the viscosity are well known from independent measurements and, in the case of helium, even better known from theory.[14] The values of the resonator's physical dimensions used in the analysis of the data were determined by dimensional measurements and were not adjusted. The values of ϵ_i and ϵ_r were adjusted slightly to improve the quality of the fit to the helium data.

The test data were acquired over an interval of three weeks and span the pressure range 0.2 MPa to 3.2 MPa (For propane, the maximum pressure was 0.75 MPa, to avoid condensation.) Additional experimental details appear in Ref. 3. At each value of the temperature and the pressure, the complex response function of the resonator was measured at 22 uniformly spaced frequencies spanning the range $\pm f/Q$ about center frequency f of the Greenspan mode. To do this, the frequency was set by a synthesizer and lock-in amplifiers were used to measure the complex voltage at the drive and detector transducers. We computed the complex ratio (detected voltage)/(drive voltage) at each frequency and fitted Eq. (36) to the results. The main fitting parameters were $D_v=\eta/\rho$ and the speed of sound c. The other fit parameters were a complex constant describing the proportionality of the complex voltage ratio to the acoustic pressure p_2 and a complex constant representing background coupling between the transducers. In some cases, a sloping background term was justified; this added an additional complex constant. The

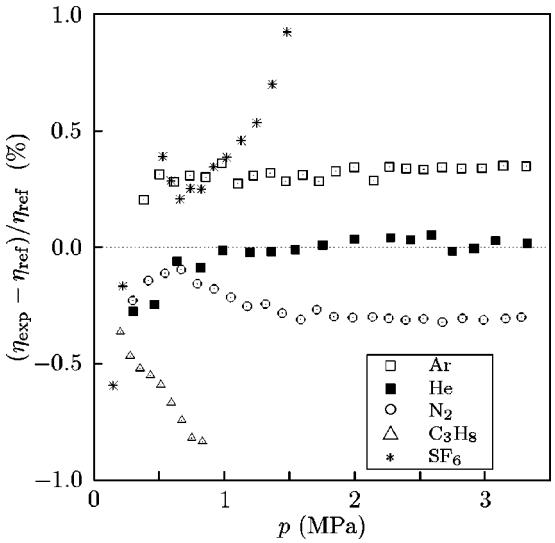


FIG. 5. Percentage differences between experimental viscosities and reference values for five gases, as functions of pressure.

fits required values of the thermal diffusivity D_t and the specific heat ratio γ. For argon, nitrogen, propane and sulfur hexafluoride, these quantities (as well as reference values of the speed of sound) were obtained from the database REFPROP;[15] for helium they were obtained from Refs. 14, 16, and 17. Typically, the standard deviation of the voltage ratio from the fit was 0.02% of the maximum voltage. If the apparatus were fully understood, a fit of this quality could determine the viscosity with an uncertainty of 0.04%.

In Fig. 5, the resulting values of $\eta(p,T)$ are compared with reference values. For this critical comparison, we rescaled the values of $\eta(p,T)$ from the database REFPROP by the values of η_0 reported by Vogel and his collaborators. (Here, η_0 is the viscosity in the limit of zero pressure at 298.15 K). In units of μPa s, the values of η_0 are nitrogen,[18] 17.782; argon,[18] 22.599; propane,[19] 8.155; sulfur hexafluoride,[20] 15.234. The reliability of the values of η_0 determined by Vogel *et al.* is confirmed by the agreement of their ratios with the ratios measured by Berg[21] (0.06% mean fractional difference) and by the small difference (0.13%) between their value of η_0 for helium (19.860)[18] and that calculated by Hurly and Moldover.[14]

The viscometer was tested over a wide frequency range. The resonance frequencies were 920±30 Hz for helium, 326±4 Hz for nitrogen, 295±5 Hz for argon, 223±7 Hz for propane, and 115±6 Hz for sulfur hexafluoride.

The argon viscosities differ from reference values by 0.4% at higher pressures. Both the helium and argon viscosities tend toward lower values at low pressures where sloping backgrounds were needed to get acceptable fits to the resonance response. Fits with constant background have the opposite trend at low pressures, tending toward higher viscosities. The viscosity of nitrogen is 0.2% below the reference values at high pressure and closer to the reference values at low pressure. The viscosities of propane and sulfur hexafluoride are close to reference values at low pressure but differ significantly at higher pressures. In these instances, we believe that the present results are more accurate than the reference data.

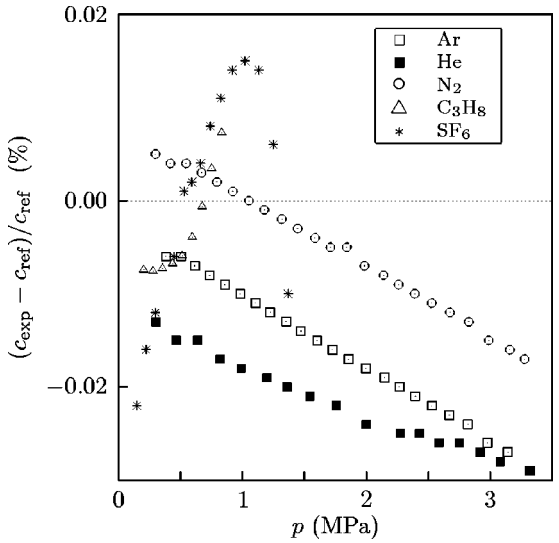


FIG. 6. Percentage differences between experimental speed of sound and reference values for five gases.

The properties of propane are representative of the polyatomic gases used in semiconductor processing. The sensitivity of the propane viscosity to the values of D_t needed for analyzing the data was tested by altering the input value of thermal conductivity by $\pm 10\%$; the corresponding changes in viscosity were $\mp 1\%$.

Figure 6 displays the deviations of the speeds of sound resulting from fitting the Greenspan viscometer data from reference values. The reference values were obtained from Ref. 14 for helium, Ref. 15 for argon and nitrogen, Ref. 22 for propane, and Ref. 23 for sulfur hexafluoride. The results for helium, argon, and nitrogen within $+0.01\%$ to -0.03% of the reference values, and have similar, approximately linear dependencies on pressure. This is surprisingly good for a resonator that was not designed to measure the speed of sound in gases. The poorer agreement of the propane and sulfur hexafluoride results is consistent with the increasing uncertainty of the reference values as condensation is approached.

As noted earlier, estimates of the elastic deformation of the viscometer have the correct order of magnitude to account for the slopes in Fig. 6.

IV. SUMMARY

A detailed acoustic model of the Greenspan viscometer has been derived. The response function, Eq. (36) depends on easily measured resonator dimensions, and the gas properties D_v, D_t, γ, ρ, and c. This equation can be fit to measurements of the resonator response function to determine the viscous diffusivity D_v and speed of sound c. The design minimizes the sensitivity of the response to the thermal effects (parameters D_t and γ), and to the fill capillary through judicious choice of the capillary location. The greatest uncertainty in the current model is in the duct end effects, which depend upon small features in the shape of the duct end that are difficult to measure. Therefore, we made modest adjustments to the numerically calculated end effects by using helium as a calibration gas. After these adjustments, the measured viscosities of four other gases agreed with reference data within $\pm 0.8\%$ over the pressure range 0.2–3.2 MPa. In future work we recommend rounding the duct ends to make the numerical calculation of end effects less sensitive to small geometric details.

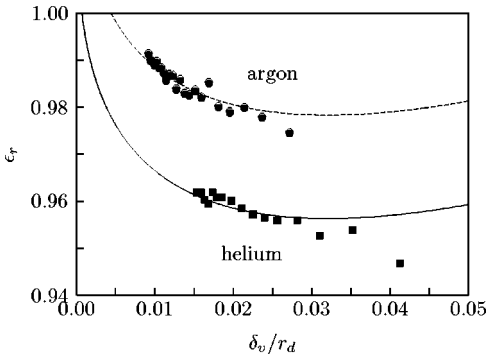


FIG. 7. Resistive end-effects parameter ϵ_r determined through fits to experimental data. ○, argon, constant background; ●, argon, sloping background; ■, helium, sloping background. The solid line is Eq. (22) with $\epsilon_{r0}=1.03$; the dashed line is Eq. (22) with $\epsilon_{r0}=1.05$.

ACKNOWLEDGMENTS

This research was supported in part by the Office of Naval Research, by NASA under Contract No. C-32088-K, and by the NIST Office of Microelectronics Programs. The authors thank Dr. J. J. Hurly and Dr. R. F. Berg for sharing data with us prior to their publication.

APPENDIX A: FITS TO DETERMINE ϵ_r

The resistive end parameter ϵ_r was determined experimentally through fits to helium and argon data. For these fits, the viscous diffusivity was assumed known and ϵ_r was a free parameter. (Reference values described previously were used for the viscous diffusivity and other thermophysical properties in the model.) Figure 7 shows that the dependence of ϵ_r on δ_v/r_d is consistent with Eq. (22). The systematic difference between He and Ar in Fig. 7 ($0.02\epsilon_r$) corresponds to the 0.4% difference between the viscosities of He and Ar in Fig. 5. The coefficient $\epsilon_{r0}=1.03$ chosen to fit the helium results was used for reduction of the data presented earlier in this paper.

APPENDIX B: VISCOMETER WITH FILL CAPILLARY IN CHAMBER

The equivalent circuit of a resonator with a fill capillary in the source chamber can be obtained from Fig. 2 by the introduction of $\mathbf{Z}_c$ in parallel with the left $\mathbf{Z}_V$. This parallel combination has an impedance

$$\mathbf{Z}_V' = \mathbf{Z}_V/(1+\mathbf{Z}_V/\mathbf{Z}_c). \tag{B1}$$

The equivalent circuit is no longer symmetric so the resonance condition is more complicated. With $\mathbf{Z}_1' = \mathbf{Z}_1 + \mathbf{Z}_{\text{end}}$, the resonance condition is

$$\mathbf{Z}_V' + \mathbf{Z}_1' + \frac{\mathbf{Z}_2(\mathbf{Z}_1' + \mathbf{Z}_V)}{\mathbf{Z}_2 + \mathbf{Z}_1' + \mathbf{Z}_V} = 0, \tag{B2}$$

which can be combined with Eq. (B1) and rearranged to get

$$\frac{\mathbf{Z}_V+\mathbf{Z}_1'}{\mathbf{Z}_V}\left[1+\frac{\mathbf{Z}_2}{\mathbf{Z}_2+\mathbf{Z}_1'+\mathbf{Z}_V}\right]=\frac{\mathbf{Z}_V}{\mathbf{Z}_V+\mathbf{Z}_c}. \tag{B3}$$

This form makes it easy to estimate the effects of the fill capillary. When the fill-duct impedance is infinite, the right-hand side vanishes and the previous resonance condition $\mathbf{Z}_V+\mathbf{Z}_1'=0$ is satisfied. When the right-hand side is small but finite, the square brackets on the left-hand side will have a magnitude near 2 because of the large magnitude of $\mathbf{Z}_2$. The first term on the left-hand side equals the negative fractional perturbation in ω^2, which accordingly is

$$-\frac{\Delta\omega^2}{\omega^2}\approx\frac{\frac{1}{2}\mathbf{Z}_V}{\mathbf{Z}_V+\mathbf{Z}_c}. \tag{B4}$$

In order to minimize the perturbation, the input impedance of the fill capillary must have a magnitude much larger than that of the chamber impedance. While it is possible to achieve this for short, capped ducts, the ducts required for filling and evacuating practical viscometers often have a length that is a significant fraction of a wavelength at the working viscometer frequency. In this case it is difficult to make the right-hand side of Eq. (B4) negligible under a wide class of operating conditions (different gases, temperatures, pressures). These considerations led Wilhelm *et al.*[3] to introduce a centered fill capillary.

The steady-state resonator response in Eq. (28), modified to account for the fill capillary, is

$$p_2=\frac{\mathbf{Z}_V\mathbf{Z}_V'\mathbf{Z}_2U_0}{(\mathbf{Z}_1'+\mathbf{Z}_V')(\mathbf{Z}_2+\mathbf{Z}_1'+\mathbf{Z}_V)+\mathbf{Z}_2(\mathbf{Z}_1'+\mathbf{Z}_V)}. \tag{B5}$$

This equation was used for analyzing data taken with the earlier versions of the viscometer.

[1] M. Greenspan and F. N. Wimenitz, "An acoustic viscometer for gases-I," NBS Report No. 2658, 1953.

[2] K. A. Gillis, J. B. Mehl, and M. R. Moldover, "Greenspan acoustic viscometer for gases," Rev. Sci. Instrum. **67**, 1850–1857 (1996).

[3] J. Wilhelm, K. A. Gillis, J. B. Mehl, and M. R. Moldover, "An improved Greenspan acoustic viscometer," Int. J. Thermophys. **21**, 983–997 (2000).

[4] M. R. Moldover, K. A. Gillis, J. J. Hurly, J. B. Mehl, and J. Wilhelm, "Acoustic measurements in gases: Applications to thermophysical properties, transport properties, and the temperature scale," in *Handbook of Elastic Properties of Solids, Fluids, and Gases* (Academic, New York, 2000), Chap. 12.

[5] M. R. Moldover, K. A. Gillis, J. J. Hurly, J. B. Mehl, and J. Wilhelm, "Acoustic measurements in gases," in *Modern Acoustical Techniques for the Measurement of Mechanical Properties*, edited by M. Levy, H. E. Bass, and R. Stern (Academic, New York, 2001), pp. 377–427.

[6] G. Kirchhoff, "Über den Einfluss der Wärmeleitung in einem Gase auf die Schallbewegung," Ann. Phys. Chem. **134**, 177 (1868); English translation "On the influence of heat conduction in a gas on sound propagation," in *Benchmark Papers in Acoustics: Physical Acoustics*, edited by R. B. Lindsay (Dowden, Hutchinson, & Ross, Stroudsburg, Pennsylvania, 1974), p. 7.

[7] I. B. Crandall, *Theory of Vibrating Systems and Sound* (Van Nostrand, New York, 1927), pp. 229–241.

[8] J. B. Mehl, "Greenspan acoustic viscometer: Numerical calculations of fields and duct end effects," J. Acoust. Soc. Am. **106**, 73–82 (1999).

[9] J. B. Mehl (unpublished).

[10] P. M. Morse and K. Uno Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Sec. 6.4.

[11] The thermal term in the corresponding expression in Refs. 2–4 was reduced by an incorrect factor of $2/\pi$. The correct formulation was, however, used for all data analysis.

[12] O. K. Mawardi, "On the propagation of sound waves in narrow conduits," J. Acoust. Soc. Am. **21**, 482–486 (1949).

[13] F. B. Daniels, "Acoustical impedance of enclosures," J. Acoust. Soc. Am. **19**, 569 (1947). (The result quoted by Daniels does not include the correction term χ_t.)

[14] J. J. Hurly and M. R. Moldover, "*Ab initio* values of the thermophysical properties of helium as standards," J. Res. Natl. Inst. Stand. Technol. **105**, 667–688 (2000).

[15] E. W. Lemmon, M. O. McLinden, and M. L. Huber, "NIST standard reference data base 23: Reference fluid thermodynamic and transport properties, version 7.0," National Institute for Standards and Technology, Standard Reference Data Program, Gaithersburg, MD, 2002.

[16] J. Kestin and W. Leidenfrost, "An absolute determination of the viscosity of eleven gases over a range of pressures," Physica (Amsterdam) **25**, 1033 (1959).

[17] J. Kestin, R. Paul, A. A. Clifford, and W. A. Wakeham, "Absolute determination of the thermal conductivity of the noble gases at room temperature up to 35 MPa," Physica A **100**, 349–369 (1980).

[18] E. Vogel, "Präzisionsmessungen des Viskositätskoeffizienten von Stickstoff und den Edelgasen zwischen Raumtemperatur und 650 K," Ber. Bunsenges. Phys. Chem. **88**, 997 (1984).

[19] E. Vogel, C. Kuechenmeister, E. Bich, and A. Laesecke, "Reference correlation of the viscosity of propane," J. Phys. Chem. Ref. Data **27**, 947 (1988).

[20] T. Strehlow and E. Vogel, "Temperature dependence and initial density dependence of the viscosity of sulfur hexafluoride," Physica A **161**, 101 (1989).

[21] R. F. Berg (private communication).

[22] J. P. M. Trusler and M. P. Zarari, "The speed of sound in gaseous propane at temperatures between 225 K and 375 K and at pressures up to 0.8 MPa," J. Chem. Thermodyn. **28**, 329–335 (1996).

[23] J. J. Hurly, D. R. Defibaugh, and M. R. Moldover, "Thermodynamic properties of sulfur hexafluoride," Int. J. Thermophys. **21**, 739–765 (2000).

Acoustic characteristics of an electrodynamic planar digital loudspeaker[a)]

Kenji Furihata,[b)] Atsushi Hayama, David K. Asano, and Takesaburo Yanagisawa
Department of Electrical and Electronic Engineering, Faculty of Engineering, Shinshu University, Wakasato 4-17-1, Nagano 380-8553, Japan

(Received 29 July 2002; revised 7 April 2003; accepted 14 April 2003)

In this paper, an electrodynamic planar loudspeaker driven by a digital signal is experimentally discussed. The digital loudspeaker consists of 22 voice coils, 11 permanent magnets, a diaphragm with streamlined sections molded in plastic, and a suspension made of handmade Japanese paper between the diaphragm and the frame. First, the acoustic responses are affected by the arrangement of the voice coils, so an asymmetric arrangement is studied. This asymmetric arrangement is designed to obtain as flat a frequency response to an analog signal as possible. This arrangement is compared with a symmetric one and results show that the flatness of the frequency response around 1 kHz and 4 kHz is improved and that the sound reproduction band is from 40 Hz to 10 kHz. Second, to evaluate the acoustic responses to a digital signal, the digital loudspeaker is driven with a pulse code modulation signal. Results show that the digital loudspeaker can reproduce pure sound with a total harmonic distortion of less than 5% from 40 Hz to 10 kHz, exceeding this value only in a narrow frequency band near 4 kHz. This digital loudspeaker was demonstrated to have good linearity over its dynamic range of 84 dB.
[DOI: 10.1121/1.1579004]

PACS numbers: 43.38.Dv, 43.38.Ja [AJZ]

I. INTRODUCTION

For decades, various kinds of audio and visual media, such as voice used in cellular phones and movies recorded on DVDs, have been digitized thanks to advancements in signal processing. For these media, digital electro-acoustic transducers that can be connected directly with many kinds of digital equipment are desirable. However, digital transducers have not advanced enough for practical use yet. The reasons are (1) it is very difficult to sample and quantize directly the mechanical vibrations from a microphone because a microphone's displacement is extremely small; (2) it is not easy to control the driving forces in a loudspeaker on a bit by bit basis.

The first digital loudspeaker was made by Flanagan.[1] This loudspeaker consisted of a round piece of electret foil, which was 46 mm in diameter. The area of the fixed electrode was divided into concentric circles and driven by 4-, 5-, and 6-bit linear PCM (pulse code modulation) signals with a constant DC voltage ($E=20$ V). After that, Hohm *et al.*[2] made a loudspeaker with a piezoelectric polymer film and experimentally studied its acoustic properties. By using a logarithmic quantized signal (A-law), higher performance and speech clarity was obtained. Also, Yanagisawa *et al.*[3] made a piezoelectric type 16-bit loudspeaker, i.e., the same number of bits used in CD and DAT recording. However, the sound pressure level was low and the frequency response was poor.

Compared with the above-mentioned loudspeakers, electro-dynamic types are expected either to be easier to make or to have a larger sound pressure level. Inanaga *et al.*[4] used D-A conversion with an electrodynamic loudspeaker, which had a conical shape with a diameter of 38 cm. Driven directly with weighted voltages, an 8-bit digital loudspeaker could achieve a frequency range of 300 Hz to 3 kHz. However, either the dynamic range or the sound reproduction band was narrow, so it did not have enough capacity for reproducing music.

In this paper, an electrodynamic planar digital loudspeaker is presented which takes into account the above limitations. In order to achieve an adequate sound pressure level and bit resolution, the loudspeaker was made with a diaphragm composed of streamlined sections and 22 voice coils. However, the acoustic responses, such as frequency and distortion, are affected by the selection of the voice coils. Therefore, in order to improve the responses, (a) the combination and (b) the arrangement of the voice coils are discussed. As for (a), pairs of voice coils are chosen by analyzing the results of each acoustic response when the loudspeaker is driven by an analog signal. In regards to (b), each bit is assigned to a voice coil based on the result of the best combination in (a).

In order to compare a digital loudspeaker with the above asymmetric arrangement to one with a symmetric arrangement, experiments were performed in which the loudspeakers were driven by a weighted discrete voltage with a maximum amplitude of 16 V_{p-p} and a resolution of 16 bits. In these experiments, the output waveform, frequency response,

[a)]Portions of this work were presented in ''Electro-dynamic planar loudspeaker driven by 16 bits digital signal and its acoustic response,'' Proceedings of WESTPRA VII, The Seventh Western Pacific Regional Acoustics Conference, Kumamoto, Japan, October 2000 (pp. 373–376), and ''Electro-dynamic type planar loudspeaker driven by 16 bits digital signal and its acoustic response,'' 17th International Congress on Acoustics, Transduction and Acoustical Devices, Rome, Italy, 7B.07.03, September 2001.

[b)]Electronic mail: kennfur@gipwc.shinshu-u.ac.jp

linearity, total harmonic distortion, distribution of sound intensity level, and directional pattern of both arrangements were evaluated.

II. ELECTRODYNAMIC PLANAR LOUDSPEAKER

In order to accomplish $D-A$ conversion at the loudspeaker directly, the binary numbers representing quantized samples of the original signal need to be decoded to PAM (pulse amplitude modulation) form and then filtered to recover the analog acoustic output. In this section, the operating principles of the digital loudspeaker and its system are described.

To increase the sound pressure level and achieve multibit resolution, an electrodynamic loudspeaker was proposed where one voice coil corresponds to one bit. However, as several voice coils were integrated into a single bobbin, a trade-off between the efficiency and resolution results because of the gap in the magnetic circuit. It is possible to achieve a resolution of up to 8 bits using a concentrically driven loudspeaker.[4,5] On the other hand, it has been reported in theory that a multiple drive panel loudspeaker has a flat frequency and omni-directional response even at high frequencies.[6] Hence, a planar type loudspeaker with multiple magnets was chosen for our digital loudspeaker.

A. Operating principles

To reduce distortion and quantization noise each sample must be represented by at least a 16-bit number giving $\pm 32\,768$ values per sample. An electrodynamic digital loudspeaker driven by a sign magnitude binary (SMB) signal is effective.[1–5] SMB is the simplest and one of the most obvious methods of encoding positive and negative numbers. The most significant bit (MSB: b_{16}) is assigned to be the sign bit. If the sign bit is 0, this means the number is positive. If the sign bit is 1, then the number is negative. The remaining 15 bits $(b_{15},b_{14},b_{13},\ldots,b_2,b_1)$ are used to represent the magnitude of the binary number in unsigned binary notation. Therefore, the formula to convert the 16-bit input signal with a length of 16 bits is

$$V_p(t) \propto (1-2b_{16})E_0 \sum_{i=1}^{15} 2^{(i-15)} b_i(t), \tag{1}$$

where V_p is the total velocity amplitude of the panel surface.

Figure 1 shows the schematic of an electrodynamic digital loudspeaker consisting of a panel, a suspension, a frame and parallel exciters with magnets and voice coils. Its operating principle is based on Fleming's left-hand rule. Provided that the velocity of the diaphragm $V_i(t)$ is controlled to be in proportion to the driving force $F_i(t)=BlI_i(t)$ of the ith exciter, the volume velocity is proportional to the input voltage $(1-2b_{16})E_0 \cdot 2^{(i-15)} b_i(t)$.

The linear dynamic properties are described by two operating equations: an electrical one and a mechanical one given by

$$(1-2b_{16})E_0 \cdot 2^{(i-15)} b_i(t) = (Z_{0i}+Z_{ei})I_i(t) + BlV_i(t), \tag{2}$$

$$0 = (Z_r+Z_p)V_i(t) - BlI_i(t). \tag{3}$$

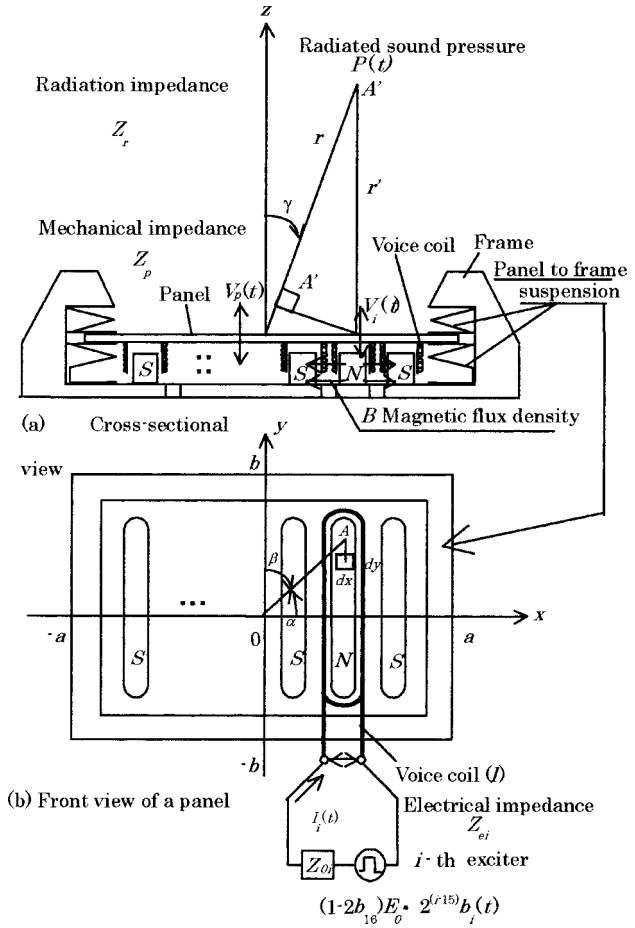


FIG. 1. Schematic of an electrodynamic planar digital loudspeaker consisting of a panel, a suspension, a frame and parallel exciters with magnets and voice coils.

The symbols used are

$(1-2b_{16})E_0 \cdot 2^{(i-15)} b_i(t)$	pulse power supply voltage applied the ith exciter (V)
E_0	voltage source (V)
Z_{0i}	internal electrical source impedance (Ω)
Z_{ei}	electrical impedance of the ith voice coil (Ω)
$I_i(t)$	current in the ith voice coil (A)
B	magnetic flux density (Wb/m^2)
l	length of the ith voice coil (m)
Bl	force factor (N/A)
$V_i(t)$	velocity amplitude of the surface of the panel driven by the ith exciter (m/s)
Z_r	radiation impedance (N s/m)
Z_p	mechanical impedance of the load including the panel and the exciter assembly (N s/m)

From the electrical Eq. (2), the current in the ith voice coil is

$$I_i(t) = \frac{(1-2b_{16})E_0 \cdot 2^{(i-15)} b_i(t)}{Z_{0i}+Z_{ei}+\dfrac{(Bl)^2}{Z_r+Z_p}}. \tag{4}$$

From the mechanical Eq. (3), the velocity of the panel driven by the ith exciter is

$$V_i(t)=\frac{Bl}{Z_r+Z_p}I_i(t). \tag{5}$$

The total velocity amplitude of the surface of the panel driven by 15 exciters, which are assumed to be a rectangular plane rigid piston mounted flush in an infinite plane baffle, is

$$V_p(t)=\sum_{i=1}^{15} V_i(t)=\frac{Bl}{Z_r+Z_p}\sum_{i=1}^{15} I_i(t). \tag{6}$$

Therefore, when the signal defined in Eq. (6) is passed through a mechanical low-pass filter, the analog drive velocity $V_p(t)$ is

$$V_p(t)=\sum_{n=-\infty}^{\infty}\frac{Bl}{Z_r+Z_p}\sum_{i=1}^{15} I_i(n\pi/\omega)\frac{\sin\omega(t-n\pi/\omega)}{\omega(t-n\pi/\omega)}, \tag{7}$$

where ω is the angular frequency.

In general, the radiation produced by the vibration of the surface of the rectangular panel [$2a\times 2b$ (m^2) as shown in Fig. 1] does not have symmetric spherical radiation patterns characteristic of a simple source. The radiation produced by the panel can, however, be found by considering them to be assemblages of simple sources whose pressure at a point is given by

$$P(t)=\left(\frac{\rho c\kappa V_{p\,\max}}{4\pi r'}\right)e^{j(\omega t-\kappa r')}, \tag{8}$$

where r' is the distance from a point A to the source, ρ is the density of air, c is the speed of sound, $\kappa=\omega/c$ is the wave number of a sound wave, and $V_{p\,\max}$ is the maximum amplitude of the analog drive velocity $V_p(t)$.

An elementary area of the surface of the panel, $dx\,dy$, can be considered to be a simple point source radiating into the infinite half-space on the upper part of the baffle. This amounts to twice the effect of the same source radiating into free space. Then

$$dP=\left(\frac{\rho c\kappa V_{p\,\max}}{2\pi r'}\right)e^{j(\omega t-\kappa r')}dx\,dy, \tag{9}$$

where r' is now the distance from point A to the $dx\,dy$ element. The total pressure at A due to the vibration of the entire piston is therefore found by integrating the above expression over the surface of the piston. Now

$$OA'=x\cos\alpha+y\cos\beta$$

$$r'=r-OA'=r-(x\cos\alpha+y\cos\beta)$$

and at a great distance from the piston, $r=r'$, so we have

$$P(t)=\frac{j\rho c\kappa V_{p\,\max}}{2\pi r}e^{j(\omega t-\kappa r)}\times\int_{-b}^{b}dy\int_{-a}^{a}e^{j\kappa(x\cos\alpha+y\cos\beta)}dx \tag{10}$$

from which

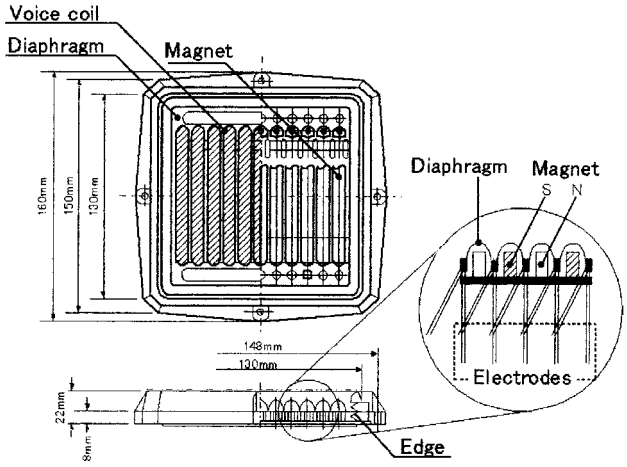


FIG. 2. Construction of our digital loudspeaker.

$$P(t)=\frac{4j\rho c\kappa abV_{p\,\max}}{2\pi r}\cdot\frac{\sin(\kappa a\cos\alpha)}{\kappa a\cos\alpha}\cdot\frac{\sin(\kappa b\cos\beta)}{\kappa b\cos\beta}e^{j(\omega t-\kappa r)}. \tag{11}$$

The radiation in the yz plane can be determined by putting $\cos\beta=\sin\gamma$, and $\sin(\kappa a\cos\alpha)/\kappa a\cos\alpha=1$. As α approaches 90°,

$$P(t)=\frac{2j\rho c\kappa abV_{p\,\max}}{\pi r}\cdot\frac{\sin(\kappa b\sin\gamma)}{\kappa b\sin\gamma}e^{j(\omega t-\kappa r)}, \tag{12}$$

where $\sin(\kappa b\sin\gamma)/\kappa b\sin\gamma$ is known as the directivity function which determines the directional characteristics of the radiation of the source. It is clear that the directional pattern becomes more pronounced at high frequencies. In other words, the greater the line dimensions of the radiator, the more pronounced the directivity will be. At the same time, minor lobes develop in addition to major lobes as the dimensions of the piston are increased.

B. Design and construction of the loudspeaker

Figure 2 shows the construction of our electrodynamic planar digital loudspeaker. The actual active part of the diaphragm is 13 cm wide, 13 cm high, and 1.1 cm deep. The diaphragm is molded plastic (total mass $m_p=0.0344$ kg, additional mass $m_{ad}=0.000\,88$ kg) and has streamlined sections to suppress the divided vibration. Two voice coils are rolled together in one section of the diaphragm, resulting in a total of 22 voice coils, each of which has a DC impedance of 4 Ω, 22 turns, and $l=4.142$ m. Also, 11 permanent magnets (each magnet is 7.5 mm wide, 87.25 mm long, and 6.0 mm thick, and each gap is 2.0 mm) are arranged under the diaphragm so that the poles of adjacent magnets are opposite each other. The suspension between the diaphragm and the frame is made of a piece of handmade Japanese paper (0.09 mm thick). The lowest resonant frequency f_0 was measured and found to be 22.7 Hz. Therefore, the stiffness s_p of the suspension is 736 N/m.

The free moving diaphragm generates sound in both the front and back of the element. The normal use of the element is to place it in an enclosure. In our test of the digital loud-

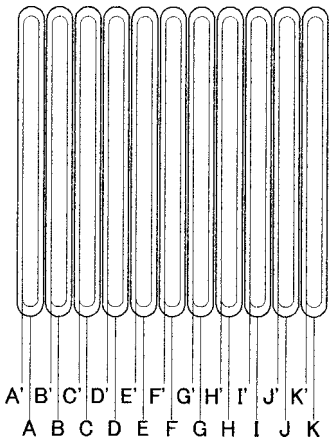


FIG. 3. The 22 voice coils in our electrodynamic loudspeaker.

speaker we used an enclosure ($V_c = 0.041\,94\ \mathrm{m}^3$) which was 28.5 cm wide, 54.5 cm high, and 27.0 cm deep. Therefore, the equivalent stiffness $s_c = \rho c^2 (ab)^2 / V_c$ is 549 N/m (20 °C).

Our planar type loudspeaker with 22 exciters has many acoustically interesting qualities, but the following technical problems must be solved:

(a) The radiating area vibrates as one unit only at lower frequencies. At higher frequencies, the radiating area exhibits more or less strong partial vibrations. Therefore, how can a flat frequency response be achieved in order to satisfy the piston motion of Eq. (6)?
(b) How can the interference be reduced within the motional electrical impedance $(Bl)^2/Z_r + Z_p$ of Eq. (4) due to motion and each exciter?
(c) In the high frequency limit, is a mechanical low-pass filter [Eq. (7)] formed as a result of the stiffness of the joint between the diaphragm and the 22 voice coils?
(d) What is the relationship between the timing precision of $D-A$ conversion and the directional responses of our electrodynamic planar digital loudspeaker?

III. ACOUSTIC RESPONSES WHEN DRIVEN BY AN ANALOG SIGNAL

The acoustic response and vibration mode are affected by the points where the diaphragm is driven by the voice coils. Therefore, the acoustic responses when driven by an analog signal are studied and improvements are discussed.

A. Measurement method

Figure 3 shows the voice coils in our electrodynamic planar loudspeaker. Each letter denotes a voice coil. Coils denoted by the same letter, such as A and A', mean that these two voice coils are rolled together in one section of the diaphragm. Experiments were carried out in an anechoic room (capacity: 60 m^3).

The fundamental test signal in the time domain is the impulse response, i.e., the output of the component when presented with a narrow rectangular voltage pulse. The impulse response carries within it a complete characterization of a component's linear performance. However, there are practical difficulties with using pulses to test loudspeakers.

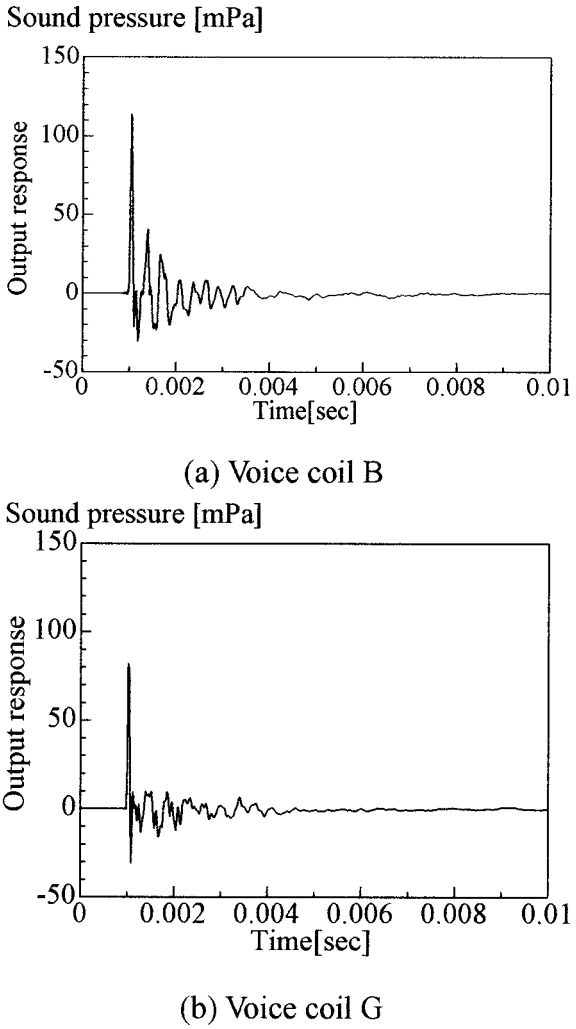


FIG. 4. Impulse responses of two voice coils (B and G).

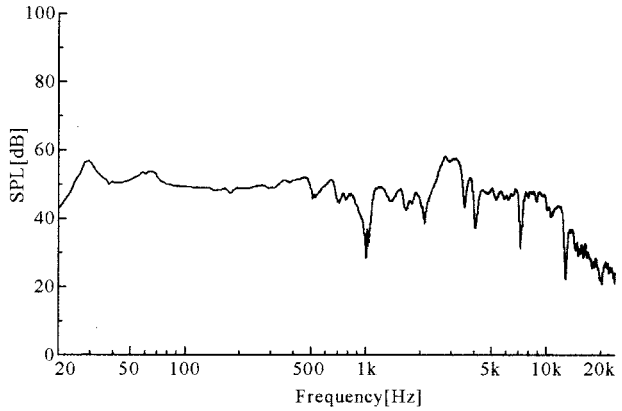
The very high dynamic range leads to a compromised signal/noise ratio, so we used a method based on maximum-length sequence test signals (sampling frequency: 48 kHz, 16th order, and average response number: 10), realized in a commercial piece of test equipment: the CTX1010 and DOCKING STATION measuring system from CORETEX Corp.

The CTX1010 and DOCKING STATION feed each voice coil a pseudorandom sequence of rectangular voltage pulses with an electrical input power of 1 W. By performing a cross correlation between the test signal and the signal picked up by a microphone with a sound level meter (RION: NL-14) at a distance of 1 m from the diaphragm, the host notebook computer (TAIWAN, R.O.C.: 7600 Series) can calculate the system's impulse response.

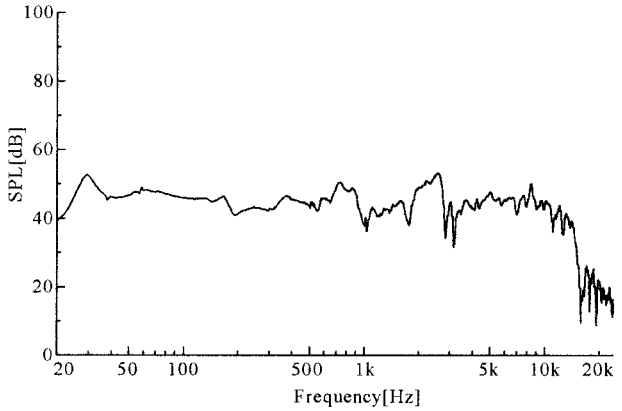
The sensitivity of a speaker is defined as the free-field sound pressure level (SPL) produced by a 1 W electrical input signal, measured at an on-axis distance of 1 m. To calibrate the SPL for the y axis of the frequency response, a pure tone with a frequency of 1 kHz was used.

B. Results

First, as an example, the relative impulse responses of coils B and G are shown in Figs. 4(a) and (b), respectively. The sharp up-spike and down-spike of each impulse re-

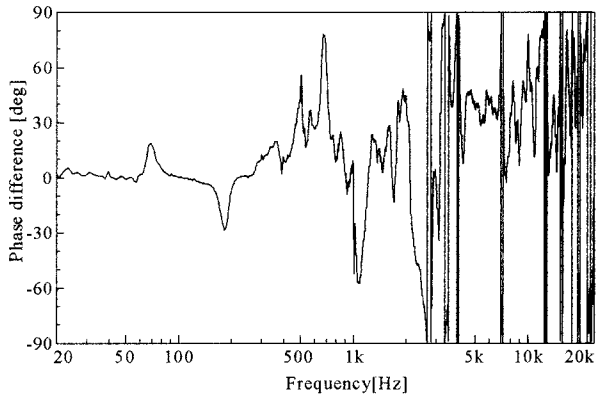


(a) Voice coil B

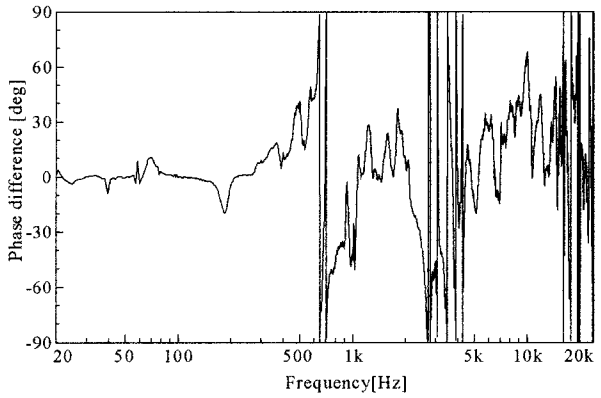


(b) Voice coil G

FIG. 5. Frequency responses of two voice coils (*B* and *G*). The vertical axis is calibrated to the free-field sound pressure level produced by a 1 W electrical input signal, measured at an on-axis distance of 1 m.



(a) Difference between voice coils B and G



(b) Difference between voice coils C and F

FIG. 6. Phase difference characteristics.

sponse is the electrodynamic planar loudspeaker's output, which is followed by lower-frequency information, a few ripples (about 0.33 ms) and reflections of the sound wave from the speaker's baffle and cabinet. The results show not only that the amplitudes are different but also that the transient responses decay slowly. Therefore, these responses may affect the cause of intersymbol interference of pulses when the loudspeaker is driven by a digital signal.

Second, the frequency responses of coils *B* and *G* are shown in Figs. 5(a) and (b), respectively. At 30 Hz a small peak appears due to the closed enclosure and the lowest resonance frequency (f_0). From 30 Hz to about 500 Hz, by acting as a piston, the diaphragm moves as a whole. Such a loudspeaker is mass-controlled over most of its pass-band. However, when it comes to higher frequencies, the radiating area exhibits more or less strong partial vibrations. Especially, at about 3 kHz, a peak is found in Fig. 5(a) corresponding to voice coil *B*, while a valley is found in (b) corresponding to voice coil *G*. A significant gap (maximum 15 dB re: 20 μPa at 1 m W) can be seen between the response levels. Similar phenomena are reported in the case of a panel loudspeaker.[6]

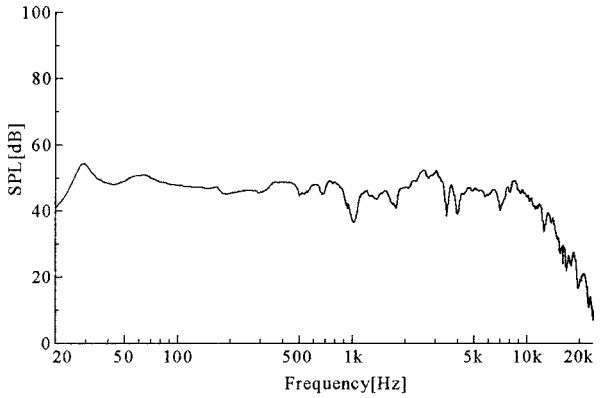
Last, Fig. 6(a) shows the phase difference between coils *B* and *G*. The phase does not deviate more than $\pm 35°$ in the region where the diaphragm acts as a piston, but there are large phase differences, especially at frequencies larger than 500 Hz. The same result (voice coils *C* and *F*) can be found in Fig. 6(b). Consequently, timing jitter and the distortion may result.

The measured sensitivity of our speaker (voice coil *F* only) is 56.6 dB at 1 kHz and 58.0 dB when a random noise signal of 2 V rms (band-limited to 16 kHz) is input. Because of the achievement of a flat frequency response from 40 Hz to 10 kHz, this result indicates our electrodynamic planar loudspeaker has a sensitivity problem in comparison with a conventional voice-coil cone speaker.

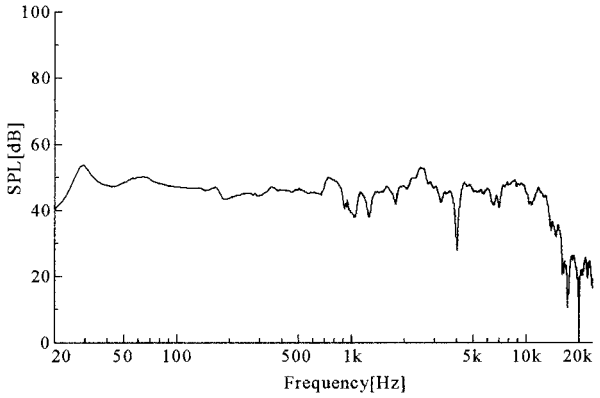
Similar acoustic responses are found in other voice coils, while coils rolled in the same section show almost the same response. The cause of these phenomena is most likely the divided vibration in the resonance or antiresonance frequencies, which vary with position.

C. Discussion

As mentioned in the preceding section, electrodynamic planar loudspeakers have some problems that may cause large distortion when driven by a digital signal because each voice coil's response is not the same. Therefore, in order to accomplish the $D-A$ conversion at the loudspeaker, it is necessary to flatten each response. One solution is to connect two voice coils in series if one coil shows a peak in its frequency response, while the other shows a dip.

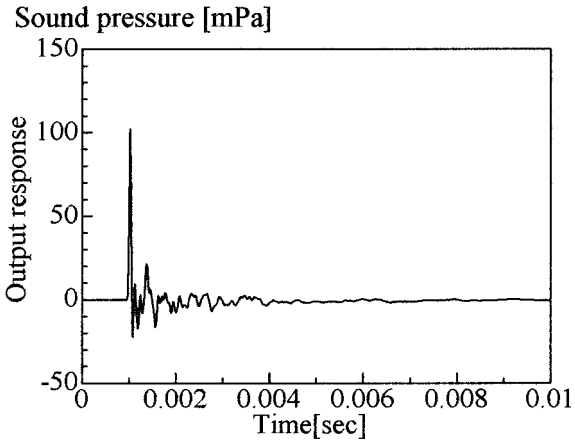


(a) Voice coils B and G

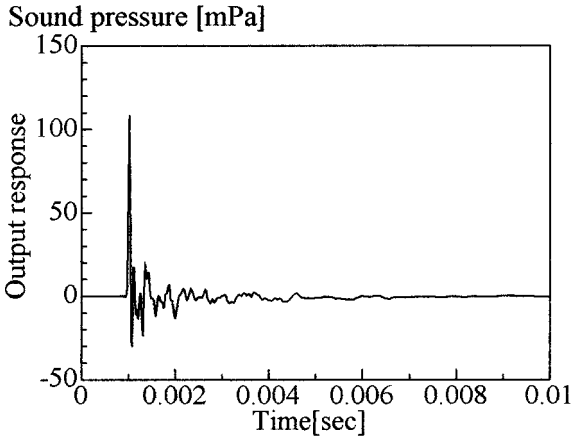


(b) Voice coils C and F

FIG. 7. Frequency response of electrodes connected in series.



(a) Voice coils B and G



(b) Voice coils C and F

FIG. 8. Impulse response of each pair of voice coils.

1. Acoustic response improvement

First, Fig. 7(a) shows the frequency response when electrodes *B* and *G* are driven simultaneously. Comparing Fig. 7(a) with Figs. 5(a) and (b), we find that the dip in the frequency response at around 3 kHz is reduced by 10 dB. The same result (in the case of the electrodes *C* and *F*) can be seen in Fig. 7(b). The results show that the loudspeaker acts as a low-pass filter with a cutoff frequency of 10 kHz and a slope of about -18 dB/octave.

Second, Fig. 8(a) shows the relative impulse response of electrodes *B* and *G* connected in series. Compared with Figs. 4(a) and (b), an improvement in transient response can be seen. The same result (in the case of electrodes *C* and *F*) can be seen in Fig. 8(b). This result is due to the enlarged effective area of vibration.

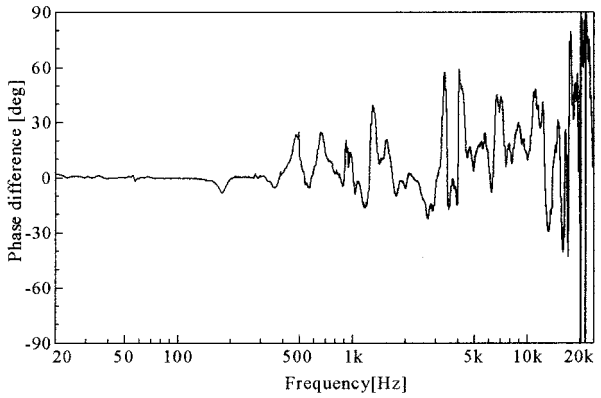
Finally, Fig. 9(a) shows the phase difference response between electrodes $B+G$ and $C+F$. The phase only deviates from $+55°$ to $-20°$ in the operational band. The same result (in the case of electrodes $E+J$ and $B+G$) can be seen in Fig. 9(b). The results show that the phase difference is diminished and a little improvement in flatness can be found at frequencies less than 10 kHz when compared with Figs. 6(a) and (b).

2. Combination of voice coils

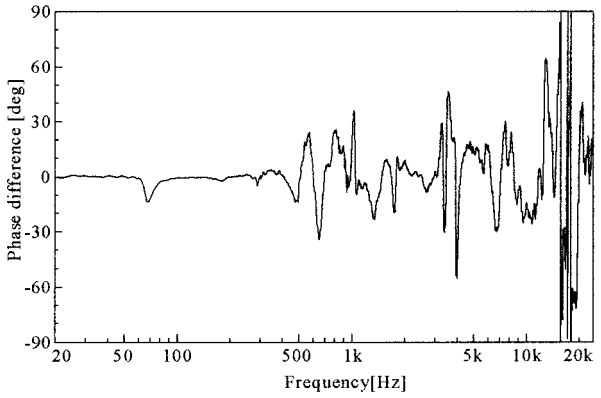
From the above results, we found that the acoustic responses can be improved by an appropriate combination of voice coils. When a 16-bit digital signal is used with an SMB code, 1 bit is used to select the direction of movement of the diaphragm and the other 15 bits are used to select the amount of diaphragm movement. Since two voice coils are required for each bit, a total of $15\times2=30$ coils are necessary. However, it is not possible at present to make 30 coils, so the combination and arrangement of the 22 voice coils in our loudspeaker are decided as follows:

(a) Two voice coils are used with each of the upper 7 bits. For each of the lower 8 bits only one coil is used.
(b) The coil positions for the upper 7 bits are chosen to make the frequency response as flat as possible.

Figure 10(b) shows the relationship between the bit numbers and voice coils that was decided according to the above methods. Coils connected in series are denoted by the same circled number. As bit numbers are not assigned symmetrically with respect to the center of the diaphragm, we call this arrangement of voice coils the ''asymmetric arrangement.'' On the other hand, the arrangement in Fig. 10(a) is called the ''symmetric arrangement'' because bit numbers are assigned symmetrically with respect to the center of the diaphragm. However, this arrangement does not follow the above-mentioned methods. Next, we discuss the difference in acoustic responses between these two arrangements.



(a) Difference between B+G and C+F



(b) Difference between E+J and B+G

FIG. 9. Phase difference characteristics.

IV. RESPONSES TO A DIGITAL SIGNAL

A. PCM driving circuit

Figure 11 shows the block diagram of the PCM driving circuit for our digital loudspeaker. The computer (NEC: LW500J/2) generates the digital audio signal, which is sent to the audio data interface (NITTOBO: AD216). Next, the digital signal is converted from a serial code into a parallel one, and then changed from Offset Binary (OB) code into SMB code in order to drive the loudspeaker. Therefore, the

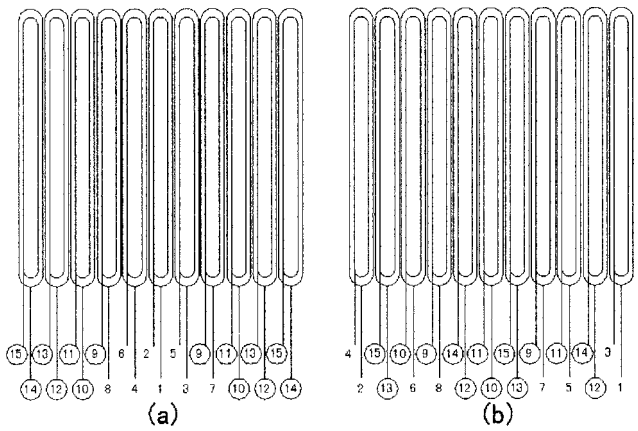


FIG. 10. The arrangement of the electrodes. (a) Symmetric arrangement. (b) Asymmetric arrangement. Circled numbers that are the same indicate a series connection.

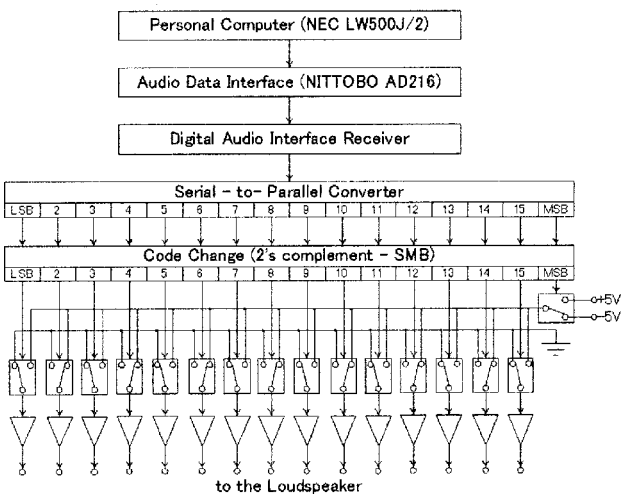


FIG. 11. Block diagram of the PCM driving circuit.

digital audio signal is changed from two levels into three, which enables the loudspeaker to vibrate in both backward and forward directions. Figure 12 shows the output waveforms of the 15th, 14th, and 13th bits of the PCM driving circuit. Table I shows the relationship between the bit number and the peak-to-peak voltage of each bit waveform.

B. Measurement method

The experiments were carried out in an anechoic room.

First, the input signal to the loudspeaker was generated by a computer (NEC: LW500J/2), which was conveyed to the PCM driving circuit. The digital audio signal had a sampling frequency of 48 kHz, and was quantized to 16 bits. Pure tones from 20 Hz to 20 kHz were used every 1/3 octave. The peak level of the input signal was $2^{15}-1 = 37\,767$ which corresponds to 0 dB. This level was decreased in 6 dB steps and the acoustic responses were measured at each step. Moreover, a rectangular tone-burst covering a fairly wide frequency band was used to evaluate the dynamic behavior of the digital speaker. The transient behavior of the loudspeaker is indicated by a change in the envelope of the burst signal. The on-axis response was measured at 10 cm from the front of the diaphragm.

Second, it is difficult to measure directly the mode pattern on the uneven diaphragm. To help identify where the

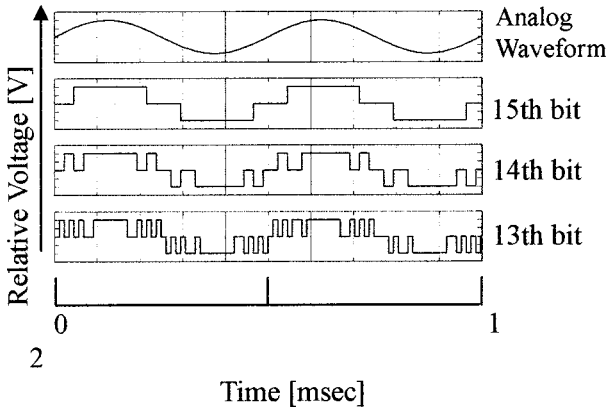


FIG. 12. Output waveforms of our PCM driving circuit.

TABLE I. Relationship between the bit number and peak-to-peak voltage.

BIT	Voltage (V_{p-p})	BIT	Voltage (V_{p-p})
1	0.001	9	0.25
2	0.002	10	0.5
3	0.004	11	1.0
4	0.008	12	2.0
5	0.015	13	4.0
6	0.031	14	8.0
7	0.062	15	16.0
8	0.125		

radiating area exhibits more or less strong partial vibrations, sound intensity distributions were measured on grids (at intervals of 13 mm) at a distance of 10 cm from the front of the diaphragm. The sound intensity analyzer used here is made by RION Co. Ltd. (sound intensity probe: SI-21, FFT analyzer: SA-74).

Third, using a microphone (RION NL-14) positioned at a distance of 1 m and at angles from 0° to 360° in 5° increments, the directional patterns of the loudspeaker were measured.

C. Results

The performance was measured using the following indices.

1. Output waveforms

Example output waveforms (pure tone, 15 bit: 0 dB, and 13 bit: −12 dB) from our electrodynamic planar digital loudspeaker with (a) the symmetric arrangement and (b) the asymmetric arrangement at 100 Hz are shown in Fig. 13. The following results were obtained.

(a) Our digital loudspeaker can reproduce pure sounds.
(b) There is a little interference between the motional and electrical components [Eq. (4)] due to the motion of the diaphragm and the exciters which drive the diaphragm.
(c) The timing precision of $D-A$ conversion on the diaphragm is good.

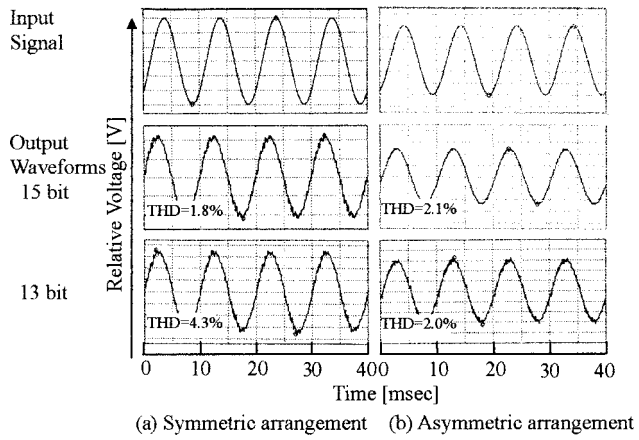


FIG. 13. Output waveforms from the electrodynamic planar digital loudspeaker at 100 Hz.

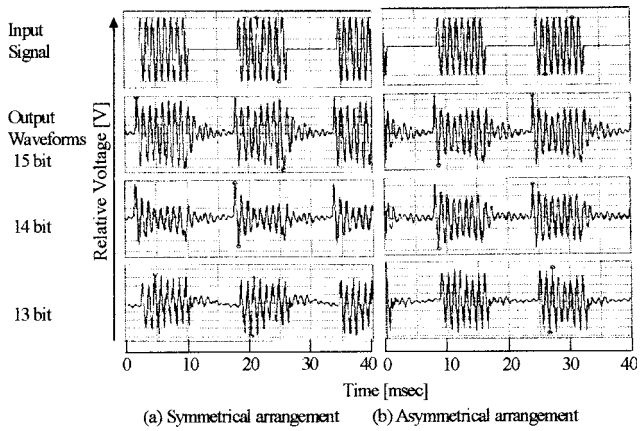


FIG. 14. Output waveforms driven by a tone burst signal at 1 kHz.

Figure 14 shows the output waveforms from the digital loudspeaker using (a) the symmetric arrangement and (b) the asymmetric arrangement driven by a tone burst signal at 1 kHz. From this figure, there is evidence of a switching transient, and seems to be about as good as is possible. Overhang (ringing after the signal is gone) is present to a small degree.

2. Frequency response

Figure 15 shows the frequency responses to 16-bit digital signals with various peak levels. We found that the asym-

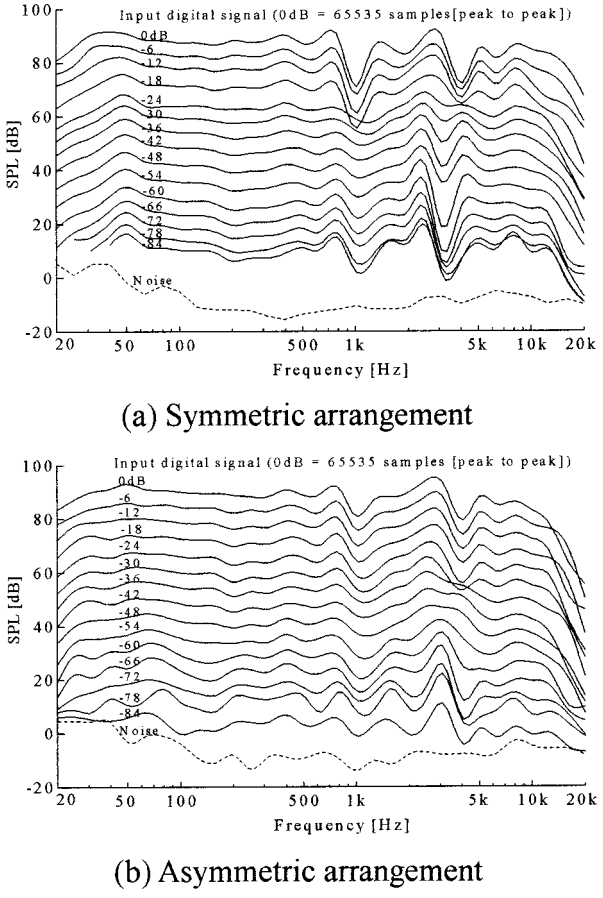


FIG. 15. Frequency responses of the digital loudspeaker at a distance of 10 cm from the center of the diaphragm. The dashed lines show the ambient noise level.

TABLE II. Relationship between SPL responses and input level from 0 dB to −84 dB.

	Symmetric arrangement			Asymmetric arrangement		
	Regression coefficient		Coefficient of	Regression coefficient		Coefficient of
Frequency	Slope	Intercept	determination	Slope	Intercept	determination
50 Hz	0.912	92.6	0.9976	1.034	92.7	0.9995
100 Hz	0.928	88.0	0.9985	1.034	91.0	0.9986
1 kHz	0.840	74.9	0.9798	0.968	82.1	0.9958
4 kHz	0.871	80.7	0.9718	0.987	83.0	0.9596
10 kHz	0.892	83.0	0.9974	0.997	84.5	0.9984

metric arrangement (b) can reduce the peak and dip at around 1 kHz and 3 kHz, and improve the flatness of the responses compared to the symmetric arrangement (a). Here the sound reproduction band is from 40 Hz to 10 kHz.

3. Linearity

Table II shows the linearity characteristics at 50 Hz, 100 Hz, 1 kHz, 4 kHz, and 10 kHz. For the asymmetric arrangement, the slopes from 50 Hz to 10 kHz are improved to be about 1.0. At the same time, because the linearity range is 84 dB, it can be said that 16-bit $D-A$ conversion is performed on the diaphragm.

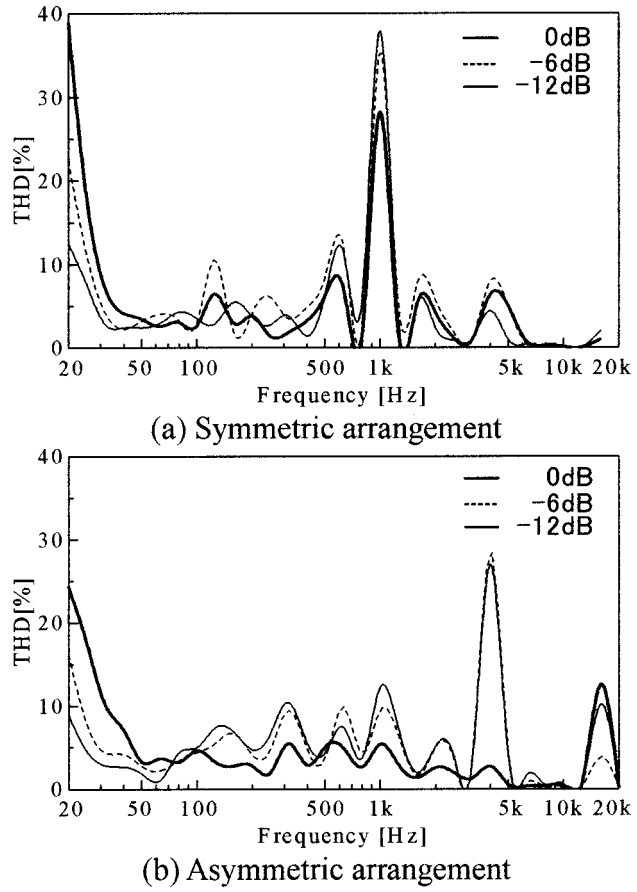


FIG. 16. Total harmonic distortion (THD) vs frequency when the input signal was changed from 0 dB to −12 dB in 6 dB steps.

4. Total harmonic distortion

THD stands for ''total harmonic distortion.'' The word ''total'' refers to the fact that the number shown represents the total of all harmonics. This total is a geometric total, formed by taking the square root of the sum of the squares of the amplitude of each of the harmonics. The THD of the digital loudspeaker using the asymmetric arrangement is about 2% at 100 Hz as shown in Fig. 13.

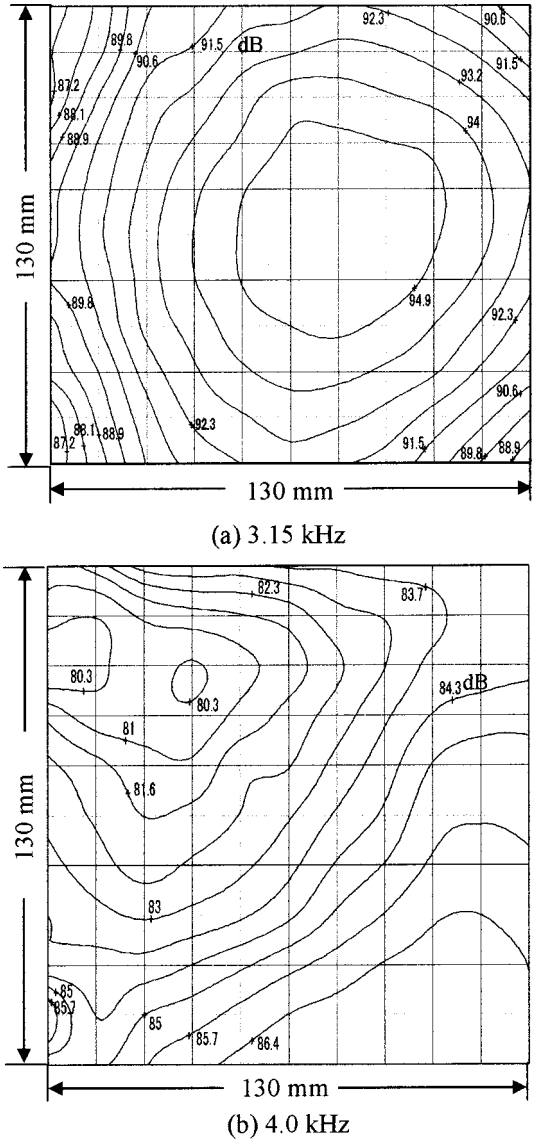


FIG. 17. Distribution of the sound intensity level (dB re: 10^{-12} W/m^2) at a distance of 10 cm from the front of the diaphragm.

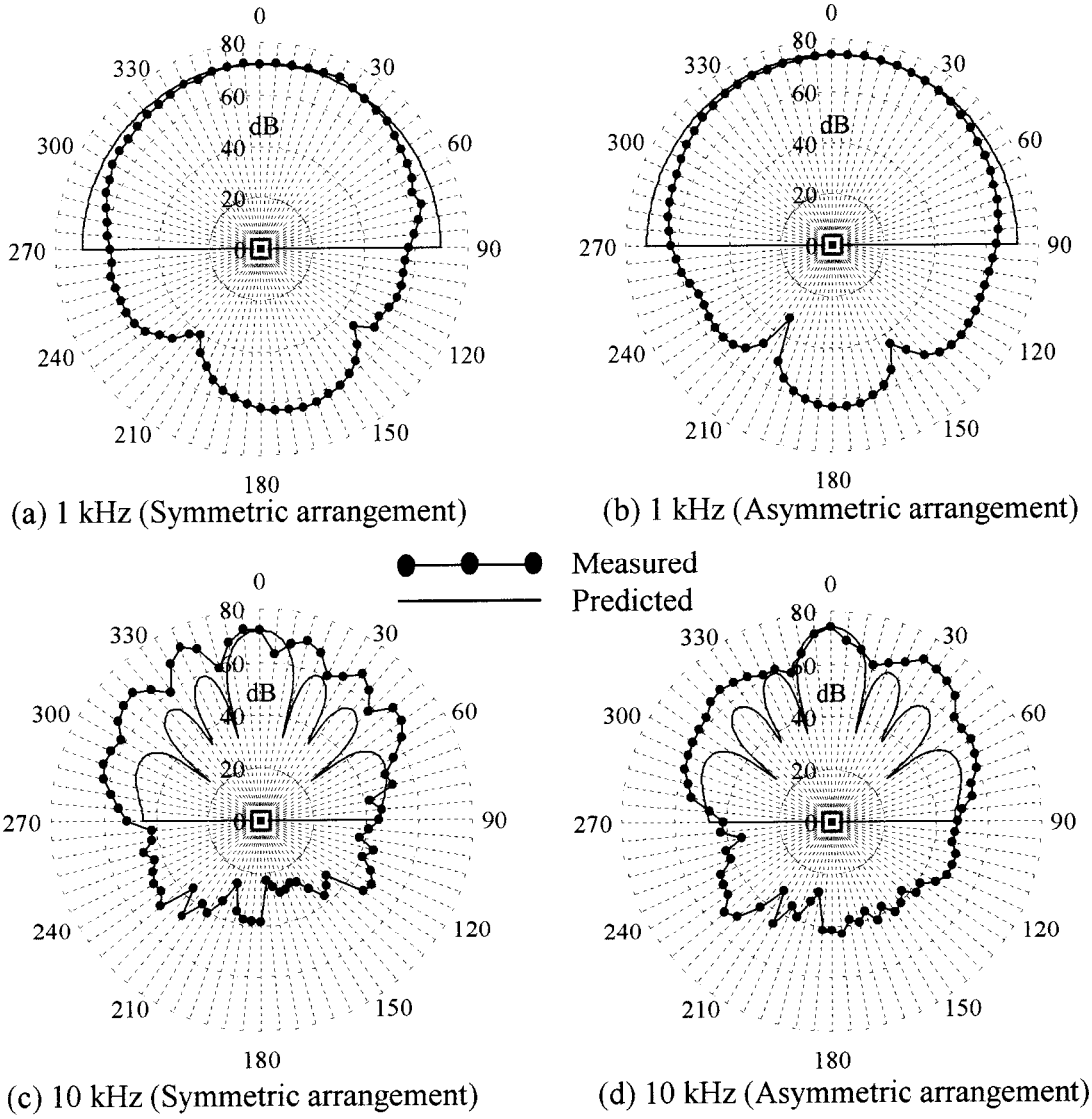


FIG. 18. Directional patterns of the loudspeaker at 1 kHz and 10 kHz. The radial scales are in SPL (dB) with the outermost circle corresponding to 80 dB *re*: 20 μPa.

Figure 16 shows the distortion factors when the input signal is changed from 0 dB to −12 dB in 6 dB steps. The reason why the THD is inferior at lower frequencies is that the voice coils separate from the magnetic circuit. The asymmetric arrangement results in an improvement in the characteristics around 1 kHz, which is relatively important for the sense of hearing. However, a new distortion at 4 kHz occurs in the asymmetric arrangement. These changes in the distortion are thought to be caused by mode transitions which depend on the arrangement of the voice coils.

5. Distribution of sound intensity level

As is evident from the contour maps of sound intensity level shown in Fig. 17, the diaphragm movement changes from (a) piston motion at 3.15 kHz to (b) divided vibration mode motion at 4 kHz. Especially, from Fig. 17(b), it can be seen that the sound intensity is highest, 86.4 dB, on the lower right surface, and lowest, 80.3 dB, on the upper left surface.

6. Directional pattern

Figure 18 shows the measured directional response (●) of the digital loudspeaker versus the theoretical data (—) based on the directivity function of Eq. (12) for a rectangular plane rigid mounted flush in an infinite plane baffle. Only theoretical data in half-space are shown because the diaphragm is embedded in the baffle. As can be seen from Fig. 18, it is clear that the directional pattern becomes more pronounced at high frequencies. It can be seen that both the measured and theoretical patterns have about six minor lobes in addition to a major lobe in half-space at 10 kHz. Because the diaphragm molded in plastic has streamlined sections as shown in Fig. 2, the measured result indicates that the asymmetric arrangement can more effectively suppress the appearance of the side lobe at 10 kHz than the predicted result based on Eq. (12).

Consequently, it can be said that the baffled-plated idealization as mentioned in the operating principles in Sec. II A is useful for our digital loudspeaker.

7. Auditory impressions

We evaluated auditory impressions with a short listening experiment. When many subjects listened to the music produced by our digital loudspeaker in a normal room (68 m^3, reverberation time of 0.8 s at 500 Hz), their auditory impressions were good for the perception of loudness, naturalness, and clearness. However, there was an impression of a little distortion peculiar to the digital speaker.

V. CONCLUSIONS

In this paper, an electrodynamic planar loudspeaker driven by a digital signal was experimentally discussed. The diaphragm was molded in plastic to have streamlined sections in order to suppress divided vibrations. Two voice coils were rolled together in one section of the diaphragm, resulting in a total of 22 voice coils. Also, 11 permanent magnets were arranged under the diaphragm so that the poles of adjacent magnets were opposite each other. The suspension between the diaphragm and the frame was made of a piece of handmade Japanese paper. The acoustic responses were affected by the arrangement of the voice coils on the diaphragm.

Therefore, first, the combination and arrangement of the voice coils were studied in order to improve the acoustic responses and achieve a 16-bit digital loudspeaker. The results are as follows:

(1) Our hand-picked pairs (asymmetric arrangement) are better than arbitrarily selecting them (symmetric arrangement).
(2) Hand-picking can improve the frequency response and suppress distortion.

Second, in order to compare the asymmetric and symmetric arrangements, experiments were performed where the digital loudspeaker was driven by a weighted discrete voltage with a maximum amplitude of 16 V_{p-p} and a resolution of 16 bits. The results are as follows:

(3) Our digital loudspeaker can reproduce pure sound. The sound reproduction band is from 40 Hz to 10 kHz.
(4) A different asymmetrical arrangement can improve the acoustic responses. Specifically, the peaks and dips in the frequency response around 1 kHz and 4 kHz are diminished and the distortion factor at 1 kHz is reduced from 28% to 2.3%.
(5) The loudspeaker has a linearity range of 84 dB in the above sound reproduction band.
(6) The baffled-plated idealization is useful to determine directional responses for the electrodynamic planar loudspeaker driven by a digital signal.

To summarize, we have constructed and demonstrated a 16-bit digital loudspeaker with good performance, wide bandwidth, and a linear dynamic range of 84 dB.

[1] J. L. Flanagan, ''Direct digital-to-analog conversion of acoustic signals,'' Bell Syst. Tech. J. **59**, 9, 1693–1719 (1980).
[2] D. Hohm, J. Franz, and R. Hoffman, ''Piezopolymerhörer als elektroakustische Digital-Analog-Wandler für linear und logarithmisch quantisierte Signale,'' Acustica **57**, 218–225 (1985).
[3] T. Yanagisawa and Y. Nakagawa, ''Piezo-electric type loudspeaker driven by 16 bits digital signal and its acoustic responses (in Japanese),'' J. Inst. Electron., Inf. Commun. Eng. **76**, 1392–1395 (1993).
[4] K. Inanaga and M. Nishimura, ''On the possibility and problems of the PCM digital loudspeaker (in Japanese),'' Technical Report of the Journal of the Institute of Electronics, Information and Communication Engineers, EA81-19, 1981, pp. 62–68.
[5] H. Uchida, T. Kusaka, K. Furihata, and T. Yanagisawa, ''Possibility of broad-band and multi-bits of electro-dynamic type digital loudspeaker (in Japanese),'' AES 9th Regional Convention, Tokyo, 1999, pp. 80–83.
[6] M. R. Bai and T. Huang, ''Development of panel loudspeaker system: Design, evaluation and enhancement,'' J. Acoust. Soc. Am. **109**, 2751–2761 (2001).

Edge waves in poroelastic plate under plane stress conditions

P. Malla Reddy[a] and M. Tajuddin
Department of Mathematics, Osmania University, Hyderabad 500007, A.P., India

(Received 28 February 2002; accepted for publication 12 November 2002)

Employing Biot's theory, the problem of edge waves in poroelastic plate under plane stress conditions is studied for both a pervious and an impervious surface. The equations for plane stress conditions are derived and discussed. The particle trajectory is obtained, which is an elliptic. The frequency equation is discussed for symmetric and antisymmetric motions and results of special interest are shown. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1569258]

PACS numbers: 43.40.At [PJR]

I. INTRODUCTION

Wave propagation in elastic media has wide applications in many branches of Physical Sciences and Engineering. A substantial amount of literature is available in Ewing *et al.*, (1957) and several detailed accounts of wave propagation are given by Kolsky (1963) and Miklowitz (1960), Kumar studied edge waves in elastic plates (Kumar, 1960).

If stress propagation is considered in plate, the behavior near the edge is different from the interior of the plate far from the edge and is comparable with the surface waves. These waves at the edge of the plate are termed as ''edge waves.'' It appears that there exist a mathematical similarity between titled problem and surface waves in a semi-infinite poroelastic solid. Accordingly, the significance of edge waves has been considerably increased in the recent past, especially in measuring velocities at which waves propagate in various parts of the earth.

Liquid saturated porous rocks are often present on and below the surface of the earth. Layers of porous solids such as sandstone or limestone saturated with ground water or oil are present in the earth crust.

Using the analytical model based on Biot's theory (Biot, 1956), a review of the work done was given by Paria (1963). A historical review of formulation of porous media theories is given by De Boer and Ehlers (1988). Taber studied deflection of poroelastic plates (Taber, 1992). Edge waves in poroelastic plates have been studied by Reddy and Tajuddin (1996). Some problems revealing interesting phenomena which characterize Biot's theory are studied by different authors (Tajuddin, 1978; Tajuddin and Sarma, 1980; Tajuddin, 1982; Tajuddin, 1984; Tajuddin and Moiz, 1984; Tajuddin, 1990; Tajuddin and Ahmed, 1991; Thomson and Willis, 1991; Jensen *et al.*, 1994; Chein and Hermann, 1996; Reddy and Tajuddin, 2000).

In the present analysis, the study of waves propagating along (parallel to) the edges of a homogeneous and isotropic thin poroelastic flat plate of infinite length is made using Biot's theory, which is in a state of plane stress. The governing equations of the plane stress problem of the said theory are not available in the literature. Thus these equations are derived from the corresponding equations of the general three dimensional problem and it is found that liquid pressure, in general, is less than that of the plane strain problem. Then these equations are employed to seek the solution of the titled problem. The equation for the particle path is obtained and it is found that every point of the plate moves in an elliptic orbit, the size and shape of which varies from one point to the other. It is seen that there occur two uncoupled families of waves for both pervious and impervious surfaces. One of these consists of motion symmetrical about midline of the plate and other one is of antisymmetrical. Frequency equations of symmetrical and antisymmetrical motions for both pervious and impervious surfaces are obtained. By ignoring liquid effects, and after rearrangement of terms, one can recover the results of purely elastic solid (Kumar, 1960). Throughout our analysis, it is shown how the results of some earlier works follow as particular cases of the more general results presented here.

The investigation is of practicable interest especially in Engineering and Geological sciences analogous to Lamb-type waves.

II. EQUATIONS FOR PLANE STRESS PROBLEM IN POROELASTIC SOLIDS

The solutions of problems in every branch of elasticity for three-dimensional bodies involves great mathematical difficulties. One can overcome these situations by making the problem to be a plane stress.

A state of stress in a plate is said to be plane if the stress vector in planes parallel to the bases is zero throughout its volume. A body is in the state of plane stress parallel to *xz* plane if stress components

$$\sigma_{xy}=\sigma_{yy}=\sigma_{yz}=\frac{\partial s}{\partial y}=0, \tag{1}$$

throughout its volume. The governing equations of plane stress problems in poroelastic solids are not readily found in literature. Therefore, following on lines of purely elastic solid, the related equations in poroelastic media are derived from the corresponding equations of the general three-dimensional problem so as to make its applications to specific boundary value problems of physical relevance.

Consider a rectangular coordinate system *Oxyz*. Let $\mathbf{u}=(u,v,w)$ and $\mathbf{U}=(U,V,W)$ be the displacement vectors of

[a]Current address: Room No. 550, Department of Electrical Engineering, Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan 106, ROC.

solid skeleton and liquid, respectively. In view of (1), the stress displacement relations of Biot (1956) yield

$$\sigma_{xx}=2N\frac{\partial u}{\partial x}+A^*e+Q^*\varepsilon,$$

$$\sigma_{xz}=N\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right),$$

$$\sigma_{zz}=2N\frac{\partial w}{\partial z}+A^*e+Q^*\varepsilon, \tag{2}$$

$$s=Q^*e+R^*\varepsilon,$$

$$\frac{\partial u}{\partial y}+\frac{\partial w}{\partial x}=0,\quad \frac{\partial u}{\partial z}+\frac{\partial w}{\partial y}=0,$$

where $e(=\operatorname{div}\mathbf{u})$ and $\varepsilon(=\operatorname{div}\mathbf{U})$ are dilatation of solid and liquid media, respectively, and A^*, Q^*, R^* are modified poroelastic constants in a plane stress problem given by

$$A^*=\frac{A(R+2N)-Q^2}{A+2Q+R+2N},$$

$$Q^*=\frac{Q(Q+2N)-AR}{A+2Q+R+2N}, \tag{3}$$

$$R^*=\frac{R(A+2N)-Q^2}{A+2Q+R+2N}.$$

In (3), A^*, Q^*, R^*, and N are all positive in sign and we infer that $Q^*<Q$, $R^*<R$ following Biot (1956). Hence, liquid pressure s in the plane stress problem is, in general, less than that of the plane strain and the general three-dimensional problems.

Setting the stress-displacement relations (2) into the equations of motion of Biot (1956), the relevant equations for plane stress problem reduce to

$$N\nabla^2u+(A^*+N)\frac{\partial e}{\partial x}+Q^*\frac{\partial \varepsilon}{\partial x}=\frac{\partial^2}{\partial t^2}(\rho_{11}u+\rho_{12}U)+b\frac{\partial}{\partial t}(u-U),$$

$$N\nabla^2w+(A^*+N)\frac{\partial e}{\partial x}+Q^*\frac{\partial \varepsilon}{\partial x}=\frac{\partial^2}{\partial t^2}(\rho_{11}w+\rho_{12}W)+b\frac{\partial}{\partial t}(w-W), \tag{4}$$

$$Q^*\frac{\partial e}{\partial x}+R^*\frac{\partial \varepsilon}{\partial x}=\frac{\partial^2}{\partial t^2}(\rho_{12}u+\rho_{22}U)-b\frac{\partial}{\partial t}(u-U),$$

$$Q^*\frac{\partial e}{\partial z}+R^*\frac{\partial \varepsilon}{\partial z}=\frac{\partial^2}{\partial t^2}(\rho_{12}w+\rho_{22}W)-b\frac{\partial}{\partial t}(w-W).$$

In (4), b is a dissipative coefficient and ρ_{ij} are mass coefficients following Biot (1956) such that the sums $\rho_{11}+\rho_{12}$ and $\rho_{12}+\rho_{22}$ represent mass of solid and liquid per unit volume of bulk material. The coefficient ρ_{12} is a mass coupling parameter between liquid and solid phases. Furthermore, the mass parameters obey the inequalities

$$\rho_{11}>0,\quad \rho_{12}\leqslant 0,\quad \rho_{22}\geqslant 0,\quad (\rho_{11}\rho_{22}-\rho_{12}^2)>0.$$

Equations (4) together with (2) have to be satisfied at every interior point of the body, and on the surface of the body, the stresses must satisfy the boundary conditions. Thus, for a plane stress problem, the solutions of a dynamical problem in poroelastic solids can be completely determined. Considering the analysis of plane time harmonic waves in a poroelastic half-space, one can see that there exists two dilatational waves and one shear wave as in Biot (1956) and its wave velocities are given by Gardner (1962) with P, Q, R replaced by P^*, Q^*, R^*, respectively. Accordingly, we see that the dilatational wave velocities will be affected, while the shear wave velocity remains the same as that of plane strain problem.

III. EDGE WAVES

Let the plane waves propagate along the edges of a poroelastic plate occupying the Cartesian space $-\infty<x<\infty$, $-\delta\leqslant y\leqslant\delta$, $-\alpha\leqslant z\leqslant\alpha$. Rectangular coordinate axes are chosen such that the x axis in the direction of propagation. Assume that exciting forces are in the central plane of symmetry of the plate and δ is small so that the problem becomes one of plane stress problems perpendicular to the y axis.

The solid displacement components $\mathbf{u}(u,w)$ which can readily be evaluated from (4), representing the plane harmonic waves travelling in the x direction are

$$\begin{aligned}u=&-ik\{A_1\cosh(pz)+B_1\sinh(pz)+A_2\cosh(qz)\\&+B_2\sinh(qz)-i\beta B_3\sinh(k\beta z)\\&-i\beta A_3\cosh(k\beta z)\}e^{i(\omega t-kx)},\\ w=&\{A_1p\sinh(pz)+B_1p\cosh(pz)+A_2q\sinh(qz)\\&+B_2q\cosh(qz)-ikB_3\cosh(k\beta z)\\&-ikA_3\sinh(k\beta z)\}e^{i(\omega t-kx)},\end{aligned} \tag{6}$$

where A_1, A_2, A_3, B_1, B_2, B_3 are all constants, ω is the frequency of the wave, k is the complex wave number and p, q, and β are

$$p=k(1-c^2V_1^{*-2})^{1/2},\quad q=k(1-c^2V_2^{*-2})^{1/2},$$

$$\beta=(1-c^2V_3^{-2})^{1/2}. \tag{7}$$

In (7), $c=\omega k^{-1}$, V_1^*, and V_2^* are velocities of dilatational waves of first and second kind, respectively, for the plane stress problem defined in Sec. II, while V_3 is the shear wave velocity (Gardner, 1962). If $k=k_1+ik_2$, then for waves to be physically realistic, one should have $k_1>0$ and $k_2\geqslant 0$. Only real parts of these are physically relevant.

By substituting the displacement components u, w in (2), one obtains the relevant stresses, which are

$$\sigma_{zz}+s=\langle A_1[2Np^2+\{(A^*+Q^*)+(Q^*+R^*)\delta_1^2\}(p^2-k^2)]\cosh(pz)+B_1[2Np^2+\{(A^*+Q^*)+(Q^*+R^*)\delta_1^2\}(p^2-k^2)]\sinh(pz)+A_2[2Nq^2+\{(A^*+Q^*)+(Q^*+R^*)\delta_2^2\}(q^2-k^2)]\cosh(qz)+B_2[2Nq^2+\{(A^*+Q^*)+(Q^*+R^*)\delta_2^2\}(q^2-k^2)]\sinh(qz)-2B_3Ni\beta k^2\sinh(k\beta z)-2A_3Ni\beta k^2\cosh(k\beta z)\rangle e^{i(\omega t-kx)}, \tag{8}$$

$$\sigma_{zx}=-Nik[2A_1p\sinh(pz)+2B_1p\cosh(pz)+2A_2q\sinh(qz)+2B_2q\cosh(qz)-B_3ik(\beta^2+1)\times\cosh(k\beta z)-A_3ik(\beta^2+1)\sinh(k\beta z)]e^{i(\omega t-kx)},$$

$$s=[A_1(Q^*+R^*\delta_1^2)(p^2-k^2)\cosh(pz)+B_1(Q^*+R^*\delta_1^2)\times(p^2-k^2)\sinh(pz)+A_2(Q^*+R^*\delta_2^2)(q^2-k^2)\cosh(qz)+B_2(Q^*+R^*\delta_2^2)(q^2-k^2)\sinh(qz)]e^{i(\omega t-kx)}, \tag{9}$$

$$\frac{\partial s}{\partial z}=[A_1(Q^*+R^*\delta_1^2)p(p^2-k^2)\sinh(pz)+B_1(Q^*+R^*\delta_1^2)p(p^2-k^2)\cosh(pz)+A_2(Q^*+R^*\delta_2^2)q(q^2-k^2)\sinh(qz)+B_2(Q^*+R^*\delta_2^2)\times q(q^2-k^2)\cosh(qz)]e^{i(\omega t-kx)}. \tag{10}$$

The terms δ_1^2 and δ_2^2 appearing in (8)–(10) are defined by

$$\delta_1^2=[(P^*R^*-Q^{*2})V_1^{*-2}-(R^*M_{11}-Q^*M_{12})]\times(R^*M_{12}-Q^*M_{22})^{-1},$$

$\delta_2^2=$ similar expression as δ_1^2 with V_1^{*-2} replaced by V_2^{*-2},

and

$$M_{11}=\rho_{11}-ib\omega^{-1},\quad M_{12}=\rho_{12}+ib\omega^{-1},\quad M_{33}=\rho_{22}-ib\omega^{-1}. \tag{11}$$

The boundary conditions to be satisfied are

$$\sigma_{zz}+s=0,\quad \sigma_{zx}=0,\quad s=0 \text{ at } z=\pm\alpha \text{ (pervious surface)}, \tag{12}$$

$$\sigma_{zz}+s=0,\quad \sigma_{zx}=0,\quad \frac{\partial s}{\partial z}=0 \text{ at } z=\pm\alpha \text{ (impervious surface)}, \tag{13}$$

where 2α is the width of the plate.

Equations (8)–(13) yield the following system of homogeneous equations for the constants A_1, B_1, A_2, B_2, B_3, and A_3:

$$A_{11}S_1^*A_1+A_{11}C_1^*B_1+A_{12}S_2^*A_2+A_{12}C_2^*B_2+A_{13}C_3^*B_3+A_{13}S_3^*A_3=0,$$

$$-A_{11}S_1^*A_1+A_{11}C_1^*B_1-A_{12}S_2^*A_2+A_{12}C_2^*B_2+A_{13}C_3^*B_3-A_{13}S_3^*A_3=0, \tag{14}$$

$$A_{21}C_1^*A_1+A_{21}S_1^*B_1+A_{22}C_2^*A_2+A_{22}S_2^*B_2-A_{23}S_3^*B_3-A_{23}C_3^*A_3=0,$$

$$A_{21}C_1^*A_1-A_{21}S_1^*B_1+A_{22}C_2^*A_2-A_{22}S_2^*B_2+A_{23}S_3^*B_3-A_{23}C_3^*A_3=0,$$

$$A_{31}C_1^*A_1+A_{31}S_1^*B_1+A_{32}C_2^*A_2+A_{32}S_2^*B_2=0,$$

$$A_{31}C_1^*A_1-A_{31}S_1^*B_1+A_{32}C_2^*A_2-A_{32}S_2^*B_2=0, \tag{15}$$

$$pA_{31}S_1^*A_1+pA_{31}C_1^*B_1+qA_{32}S_2^*A_2+qA_{32}C_2^*B_2=0,$$

$$-pA_{31}S_1^*A_1+pA_{31}C_1^*B_1-qA_{32}S_2^*A_2+qA_{32}C_2^*B_2=0. \tag{16}$$

Equations (14) and (15) correspond to a pervious surface, while Eqs. (14) and (16) correspond to an impervious surface. In (14)–(16), we have

$$A_{11}=2Nikp,\quad A_{12}=2Nikq,\quad A_{13}=k^2(\beta^2+1),$$

$$A_{12}=2Np^2+\{(A^*+Q^*)+(Q^*+R^*)\delta_1^2\}(p^2-k^2),$$

$$A_{23}=2Ni\beta k^2,$$

$$A_{31}=(Q^*+R^*\delta_1^2)(p^2-k^2),$$

A_{22}, $A_{32}=$ similar expressions as A_{21}, A_{31} with p^2 and δ_1^2 replaced by q^2 and δ_2^2, respectively,

and

$$C_1^*=\cosh(p\alpha),\quad C_2^*=\cosh(q\alpha),\quad C_3^*=\cosh(k\beta\alpha),\quad S_1^*=\sinh(p\alpha),\quad S_2^*=\sinh(q\alpha),\quad S_3^*=\sinh(k\beta\alpha). \tag{17}$$

Combining and subtracting first and second equations of (14), third and fourth equations of (14) and the equations of (15), one obtains the uncoupled systems of homogeneous equations in A_1, A_2, A_3 and B_1, B_2, B_3 for a pervious surface, namely,

$$A_{11}S_1^*A_1+A_{12}S_2^*A_2+A_{13}S_3^*A_3=0, \tag{18}$$

$$A_{21}C_1^*A_1+A_{22}C_2^*A_2-A_{23}C_3^*A_3=0, \tag{19}$$

$$A_{31}C_1^*A_1+A_{32}C_2^*A_2=0, \tag{20}$$

and

$$A_{11}C_1^*B_1+A_{12}C_2^*B_2+A_{13}C_3^*B_3=0, \tag{21}$$

$$A_{21}S_1^*B_1+A_{22}S_2^*B_2-A_{23}S_3^*B_3=0, \tag{22}$$

$$A_{31}S_1^*B_1+A_{32}S_2^*B_2=0. \tag{23}$$

Arguing on similar lines for an impervious surface, we obtain two sets of equations, of which one is

$$pA_{31}S_1^*A_1+qA_{32}S_2^*A_2=0, \tag{24}$$

in addition to (18) and (19). The other set of equations is (21), (22), and

$$pA_{31}C_1^*B_1+qA_{32}C_2^*B_2=0. \tag{25}$$

From Eqs. (18)–(25), we see that all the 'A's and all the 'B's are linked together and 'A's and 'B's are unlinked. From Eqs. (18)–(23) for a pervious surface, one obtains

$$\begin{aligned} A_3A_1^{-1}&=-(A_{11}S_1^*-A_{12}S_2^*A_0)/A_{13}S_3^* \\ &=(A_{21}C_1^*-A_{22}C_2^*A_0)/A_{23}C_3^*, \end{aligned} \tag{26}$$

where

$$A_0=-A_2/A_1=A_{31}C_1^*/A_{32}C_2^*. \tag{27}$$

Also

$$\begin{aligned} B_3B_1^{-1}&=(A_{12}C_2^*B_0-A_{11}C_1^*)/A_{13}C_3^* \\ &=(A_{21}S_1^*-A_{22}S_2^*B_0)/A_{23}S_3^*, \end{aligned} \tag{28}$$

where

$$B_0=-B_2/B_1=A_{31}S_1^*/A_{32}S_2^*. \tag{29}$$

With the aid of Eqs. (26)–(29), Eqs. (6) and (9) yield the following expressions for displacement u, w, and the liquid pressure s:

$$\begin{aligned} u&=-[A_1f_1(z)+B_1g_1(z)]e^{i(\omega t-kx)}, \\ w&=[A_1f_2(z)+B_1g_2(z)]e^{i(\omega t-kx)}, \\ s&=[A_1f_3(z)+B_1g_3(z)]e^{i(\omega t-kx)}, \end{aligned} \tag{30}$$

where

$$\begin{aligned} f_1(z)&=ik\cosh(pz)-ikA_{31}A_{32}^{-1}C_1^*C_2^{*-1}\cosh(qz) \\ &\quad+k\beta(A_{21}C_1^*-A_{22}C_2^*A_0)A_{23}^{-1}C_3^{*-1}\cosh(k\beta z), \\ g_1(z)&=ik\sinh(pz)-ikA_{31}A_{32}^{-1}S_1^*S_2^{*-1}\sinh(qz) \\ &\quad+k\beta(A_{21}S_1^*-A_{22}S_2^*B_0)A_{23}^{-1}S_3^{*-1}\sinh(k\beta z), \\ f_2(z)&=p\sinh(pz)-A_0q\sinh(qz)-ik(A_{21}C_1^* \\ &\quad-A_{22}C_2^*A_0)A_{23}^{-1}C_3^{*-1}\sinh(k\beta z), \\ g_2(z)&=p\cosh(pz)-B_0q\cosh(qz)-ik(A_{21}S_1^* \\ &\quad-A_{22}S_2^*B_0)A_{23}^{-1}S_3^{*-1}\cosh(k\beta z), \\ f_3(z)&=A_{31}\cosh(pz)-A_{33}C_1^*C_2^{*-1}\cosh(qz), \\ g_3(z)&=A_{31}\sinh(pz)-A_{33}S_1^*S_2^{*-1}\sinh(qz). \end{aligned} \tag{31}$$

It is clear from Eqs. (30) that for the waves of desired type to exist, at least one of the constants A_1 and B_1 should be nonzero. If $B_1=0$, the solutions (30) represent waves consisting of symmetrical motions with respect to the x axis and if $A_1=0$, the waves consists of antisymmetrical motions. In general, solution (30) represent a superposition of two uncoupled families of waves, of which one consists of only symmetrical motion and other antisymmetrical motion with respect to the x axis.

In order that u, w, and s are to be purely real, taking the real parts on the right-hand side of Eqs. (30), one obtains

$$\begin{aligned} u&=[F_1^*\cos(\omega t-k_1x)-F_2^*\sin(\omega t-k_1x)]e^{k_2x}, \\ w&=[G_1^*\cos(\omega t-k_1x)-G_2^*\sin(\omega t-k_1x)]e^{k_2x}, \\ s&=[I_1\cos(\omega t-k_1x)-I_2\sin(\omega t-k_1x)]e^{k_2x}, \end{aligned} \tag{32}$$

where F_1^*, G_1^*, I_1 and F_2^*, G_2^*, I_2 are the real and imaginary parts of the functions $F(z)$, $G(z)$, and $I(z)$, respectively, given by

$$\begin{aligned} F(z)&=A_1f_1(z)+B_1g_1(z), \\ G(z)&=A_1f_2(z)+B_1g_2(z), \\ I(z)&=A_1f_3(z)+B_1g_3(z). \end{aligned} \tag{33}$$

Eliminating the trigonometric functions from the first two equations of (32), one can arrive at the following equation connecting u and w:

$$(G_2^*u-F_2^*w)^2+(G_1^*u-F_1^*w)^2=(F_1^*G_2^*-F_2^*G_1^*)^2e^{2k_2x}. \tag{34}$$

Equation (34) represents an ellipse in the plane of the plate. Thus, every point of the plate moves in an elliptic orbit. It is clear that the size and shape of the orbit varies from one point to another and it increases with the advancement of the waves. A similar result holds good even for an impervious surface.

IV. FREQUENCY EQUATION

We now proceed to obtain the frequency equations of symmetrical, antisymmetrical motions each for a pervious and an impervious surface. We shall discuss each case independently. In addition, it is convenient to introduce dimensionless quantities as follows:

$$\begin{aligned} &a_1=PH^{-1},\quad a_2=QH^{-1},\quad a_3=RH^{-1},\quad a_4=NH^{-1}, \\ &d_1=\rho_{11}\rho^{-1},\quad d_2=\rho_{12}\rho^{-1},\quad d_3=\rho_{22}\rho^{-1}, \\ &m=CC_0^{-1}, \\ &\tilde{x}=(V_0V_1^{*-1})^2,\quad \tilde{y}=(V_0V_2^{*-1})^2,\quad z=(V_0V_3^{-1})^2. \end{aligned} \tag{35}$$

In Eqs. (35), V_0, C_0 are reference velocities ($V_0^2=H\rho^{-1}, C_0^2=N\rho^{-1}$) and $\rho=\rho_{11}+2\rho_{12}+\rho_{22}$, $H=P+2Q+R$.

A. Symmetrical motions

Equations (18)–(20) correspond to symmetrical motion for a pervious surface. A nontrivial solution can be obtained if the determinant of coefficients vanishes. Thus one obtains the frequency equations for symmetrical motion in case of a pervious surface

$$|B_{ij}|=0\quad (i,j=1,2,3). \tag{36}$$

In Eq. (36), the elements B_{ij} are

$$\begin{aligned} &B_{11}=2D_{41}\tanh(p\alpha),\quad B_{12}=2D_{42}\tanh(q\alpha), \\ &B_{13}=(D_{43}^2+1)\tanh(k\beta\alpha), \\ &B_{21}=D_{21},\quad B_{22}=D_{22},\quad B_{23}=2a_4D_{43}, \\ &B_{31}=D_{31},\quad B_{32}=D_{32},\quad B_{33}=0, \end{aligned} \tag{37}$$

where

$$p\alpha = k_1\alpha D_{41}, \quad q\alpha = k_1\alpha D_{42}, \quad k_1\beta\alpha = k_1\alpha D_{43},$$

$$D_{21} = 2a_4 D_{41}^2 + [D_{11} + D_{12} + (D_{12} + D_{13})\xi_1^2](D_{41}^2 - 1),$$

$$D_{31} = (D_{12} + D_{13}\xi_1^2)(D_{41}^2 - 1),$$

where

$$\xi_1^2 = \frac{[(D_{11} - 2a_4)D_{13} - D_{12}^2]\tilde{x} - (D_{13}d_1 - D_{12}d_2)}{D_{13}d_2 - D_{12}d_3},$$

$$D_{11} = (a_1 - 2a_4)(a_3 + 2a_4) - a_2^2,$$

$$D_{12} = a_2(a_2 + 2a_4) - a_3(a_1 - 2a_4),$$

$$D_{13} = a_1 a_3 - a_2^2,$$

$$D_{41} = (1 - m_2\tilde{x}a_4)^{1/2}, \tag{38}$$

D_{22}, D_{32} = similar expressions as D_{21}, D_{31} with $\tilde{x}, \xi_1^2$ replaced by $\tilde{y}, \xi_2^2$, respectively,

D_{42}, D_{43} = similar expressions as D_{41} with $\tilde{x}$ replaced by $\tilde{y}, \tilde{z}$, respectively,

ξ_2^2 = similar expression as ξ_1^2 with $\tilde{x}$ replaced by $\tilde{y}$.

Arguing on similar lines, the frequency equation in the case of an impervious surface becomes

$$|C_{ij}| = 0, \ (i,j = 1,2,3), \tag{39}$$

where

$$C_{11} = B_{11}, \quad C_{12} = B_{12}, \quad C_{13} = B_{13},$$

$$C_{21} = B_{21}, \quad C_{22} = B_{22}, \quad C_{23} = B_{23}, \tag{40}$$

$$C_{31} = B_{31}D_{41}\tanh(p\alpha), \quad C_{32} = B_{32}D_{42}\tanh(q\alpha), \quad C_{33} = 0.$$

In Eq. (40), B_{ij}, D_{41}, and D_{42} are defined in (37) and (38).

In view of Eqs. (38), Eqs. (36) and (39) are implicit relations connecting m and $k_1\alpha$. Each of these equations are transcendental and complex valued, hence real roots may not exit. These equations have been reduced to a real valued real equation by assuming that the water or kerosene saturated sandstone is nondissipative. Accordingly an assumption is made to solve frequency equations numerically to obtain the velocity of propagation. The phase velocity is then simply ωk_1^{-1}. These equations yield infinitely many roots for m in terms of $k_1\alpha$, each root corresponding to a mode of vibration. The symmetrical motion is obviously dispersive in nature. Also, following as in Eringen and Suhubi (1975), Eqs. (36) and (39) are identified as frequency equations of compressional waves in a thick poroelastic plate of thickness 2α for a pervious and an impervious surface, respectively. Setting liquid effects to vanish in Eqs. (36), the frequency equation obtained by Kumar (1960) for a purely elastic solid is recovered as special case, wherein physical significance was not given.

Now we shall discuss two special cases, when $k_1\alpha \ll 1$ and $k_1\alpha \gg 1$.

Case (i) when $k_1\alpha \ll 1$: In this case all the hyperbolic tangent functions approach their respective angles, and then from Eq. (36) for a pervious surface, one obtains

$$|E_{ij}| = 0, \quad (i,j = 1,2,3), \tag{41}$$

where

$$E_{11} = 2D_{41}^2, \quad E_{12} = 2D_{42}^2, \quad E_{13} = (D_{43}^2 + 1)D_{43},$$

$$E_{21} = D_{21}, \quad E_{22} = D_{22}, \quad E_{23} = 2a_4 D_{43}, \tag{42}$$

$$E_{31} = D_{31}, \quad E_{32} = D_{32}, \quad E_{33} = 0.$$

In Eq. (42), D_{ij} are defined in (38).

Similarly, for an impervious surface, one obtains

$$|F_{ij}| = 0 \quad (i,j = 1,2,3), \tag{43}$$

where

$$F_{11} = E_{11}, \quad F_{12} = E_{12}, \quad F_{13} = E_{13},$$

$$F_{21} = E_{21}, \quad F_{22} = E_{22}, \quad F_{23} = E_{23}, \tag{44}$$

$$F_{31} = D_{31}D_{41}^2, \quad F_{32} = D_{32}D_{42}^2, \quad F_{33} = 0.$$

In (44), E_{ij} are defined in (42) and D_{ij} are defined in (38). It is seen that both Eqs. (41) and (43) are nondispersive and correspond to a frequency equation of thin poroelastic bar for a pervious and an impervious surface, respectively. In the limiting case when $k_1\alpha$ approaches zero, it is readily verified that there exists four roots for dimensionless phase velocity m, namely 0, $\tilde{x}$, $\tilde{y}$, z in turn determine the phase velocity c to be 0, V_1^*, V_2^*, V_3 examined by Reddy and Tajuddin (1996). From (41) and (43), it is clear that nature of the surface influences the symmetric motion when $k_1\alpha \ll 1$.

Case (ii) when $k_1\alpha \gg 1$: In this case all the hyperbolic tangent functions tend to unity, and then from (36), one obtains

$$|G_{ij}| = 0 \quad (i,j = 1,2,3), \tag{45}$$

where

$$G_{11} = 2D_{41}, \quad G_{12} = 2D_{42}, \quad G_{13} = D_{43}^2 + 1,$$

$$G_{21} = D_{21}, \quad G_{22} = D_{22}, \quad G_{23} = 2a_4 D_{43}, \tag{46}$$

$$G_{31} = D_{31}, \quad G_{32} = D_{32}, \quad G_{33} = 0.$$

In (46), D_{ij} are defined in (38).

Similarly, in case of an impervious surface, from (39) one obtains

$$|H_{ij}| = 0 \quad (i,j = 1,2,3), \tag{47}$$

where

$$H_{11} = G_{11}, \quad H_{12} = G_{12}, \quad H_{13} = G_{13},$$

$$H_{21} = G_{21}, \quad H_{22} = G_{22}, \quad H_{23} = G_{23}, \tag{48}$$

$$H_{31} = D_{31}D_{41}, \quad H_{32} = D_{32}D_{42}, \quad H_{33} = 0.$$

In (48), G_{ij} are defined in (46) and D_{ij} are defined in (38). It is seen that frequency equations (45) and (47) are nondispersive and are analogous to Rayleigh waves in semi-infinite

TABLE I. Two types of poroelastic materials are considered to carry out computational work. Material 1 is sandstone saturated with kerosene. Material 2 is sandstone saturated with water.

	a_1	a_2	a_3	a_4	d_1	d_2	d_3	$\tilde{x}$	$\tilde{y}$	$\tilde{z}$
Material 1	0.843	0.065	0.027	0.234	0.901	−0.001	0.1	4.869	0.998	3.85
Material 2	0.96	0.006	0.0289	0.412	0.876	0	0.124	4.2977	0.912	2.126

poroelastic solid for a pervious and an impervious surface, respectively, examined by Tajuddin (1984).

B. Antisymmetrical motions

Equations (21)–(23) correspond to antisymmetrical motions for a pervious surface. Proceeding on the lines of symmetrical motions, one obtains the nondimensional frequency equation for antisymmetrical motions in case of a pervious surface to be

$$|I_{ij}|=0 \quad (i,j=1,2,3), \tag{49}$$

where

$$\begin{aligned} &I_{11}=2D_{41}, \quad I_{12}=2D_{42}, \quad I_{13}=D_{43}^2+1, \\ &I_{21}=D_{21}\tanh(p\alpha), \quad I_{22}=D_{22}\tanh(p\alpha), \\ &I_{23}=2a_4D_{43}\tanh(k_1\beta\alpha), \\ &I_{31}=D_{31}\tanh(p\alpha), \quad I_{32}=D_{32}\tanh(q\alpha), \quad I_{33}=0. \end{aligned} \tag{50}$$

In (50), D_{ij} are defined in (38).

Similarly, the frequency equation in dimensionless form for an impervious surface is given by

$$|J_{ij}|=0 \quad (i,j=1,2,3), \tag{51}$$

where

$$\begin{aligned} &J_{11}=I_{11}, \quad J_{12}=I_{12}, \quad J_{13}=I_{13}, \\ &J_{21}=I_{21}, \quad J_{22}=I_{22}, \quad J_{23}=I_{23}, \\ &J_{31}=D_{31}D_{41}, \quad J_{32}=D_{32}D_{42}, \quad J_{33}=0. \end{aligned} \tag{52}$$

In (52), I_{ij} are defined in (50) and D_{ij} are defined in (38). It is seen that frequency equations (49) and (51) which connect m and $k_1\alpha$ are dispersive in nature and are identified as the frequency equations of flexural waves in a thick poroelastic plate of thickness 2α for a pervious and an imperious surface, respectively, following as in Eringen and Suhubi (1975). Setting liquid effects to vanish in (49), the results of a purely elastic solid is obtained as a special case considered by Kumar (1960).

Now we shall discuss the two limiting cases when $k_1\alpha \ll 1$ and $k_1\alpha \gg 1$.

Case (I) when $k_1\alpha \ll 1$: Unlike that of symmetrical motion both Eqs. (49) and (51) yield

$$|K_{ij}|=0 \quad (i,j=1,2,3), \tag{53}$$

where

$$\begin{aligned} &K_{11}=2D_{41}, \quad K_{12}=2D_{42}, \quad K_{13}=D_{43}^2+1. \\ &K_{21}=D_{21}D_{41}, \quad K_{22}=D_{22}D_{42}, \quad K_{23}=2a_4D_{43}^2, \\ &K_{31}=D_{31}D_{41}, \quad K_{32}=D_{32}D_{42}, \quad K_{33}=0. \end{aligned} \tag{54}$$

In (54), D_{ij} are defined in (38).

Unlike the case of symmetric motion, it is clear that the nature of the surface does not influence the frequency equation for antisymmetrical motion. In the limiting case when $k_1\alpha$ approaches zero, the roots of the frequency equation (53) are $m=0$, $\tilde{x}$, $\tilde{y}$, z, which in turn gives $c=0$, V_1^*, V_2^*, V_3, examined by Reddy and Tajuddin (1996).

Case (ii) when $k_1\alpha \gg 1$: In this case, the frequency equation for a pervious surface given by (49) reduces to

$$|L_{ij}|=0 \quad (i,j=1,2,3), \tag{55}$$

where

$$\begin{aligned} &L_{11}=2D_{41}, \quad L_{12}=2D_{42}, \quad L_{13}=D_{43}^2+1, \\ &L_{21}=D_{21}, \quad L_{22}=D_{22}, \quad L_{23}=2a_4D_{43}, \\ &L_{31}=D_{31}, \quad L_{32}=D_{32}, \quad L_{33}=0. \end{aligned} \tag{56}$$

In (56), D_{ij} are defined in (38).

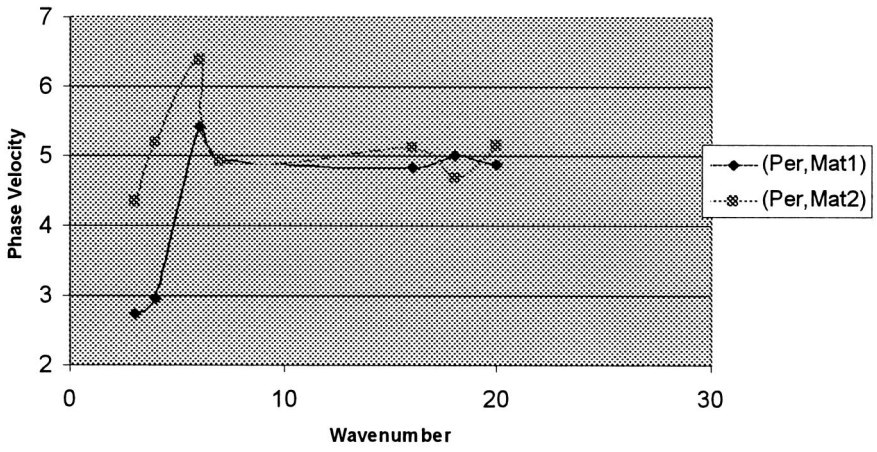


FIG. 1. Phase velocity as a function of wave number (symmetric motion, pervious surface).

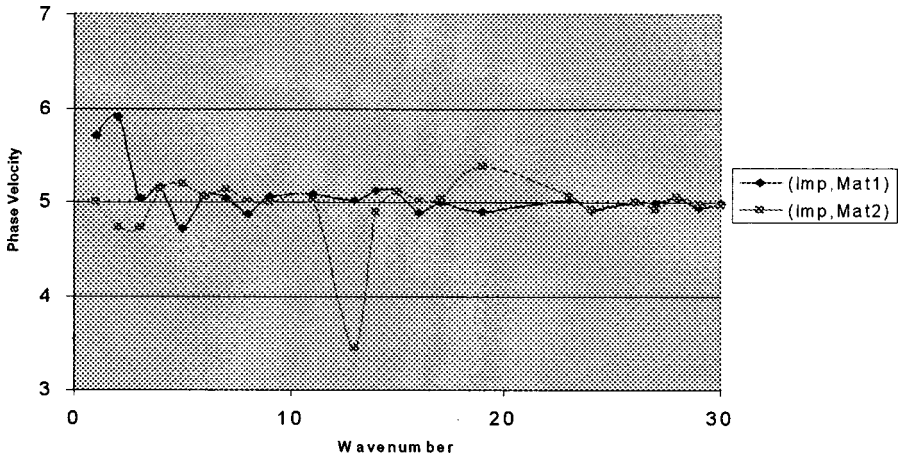


FIG. 2. Phase velocity as a function of wave number (symmetric motion, impervious surface).

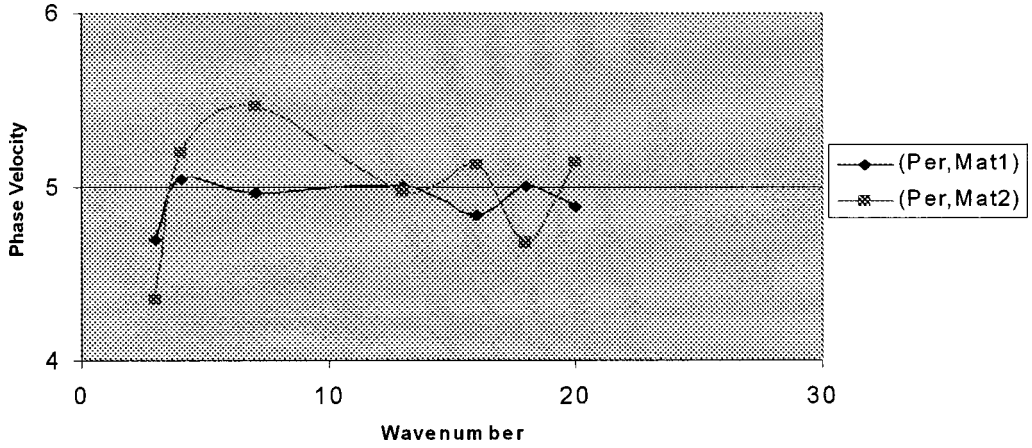


FIG. 3. Phase velocity as a function of wave number (antisymmetric motion, pervious surface).

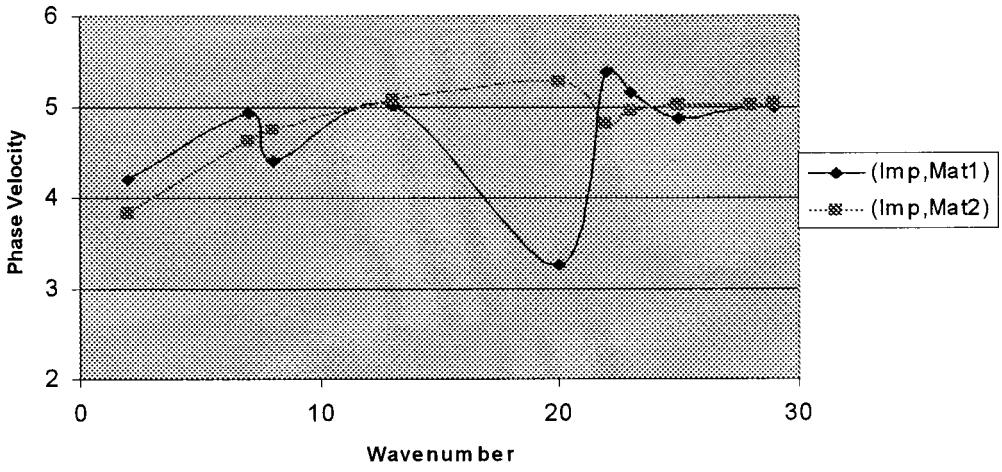


FIG. 4. Phase velocity as a function of wave number (antisymmetric motion, impervious surface).

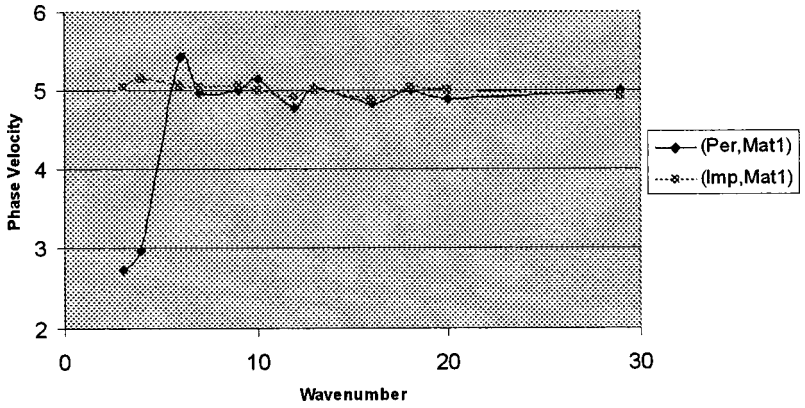


FIG. 5. Phase velocity as a function of wave number (symmetric motion, material 1).

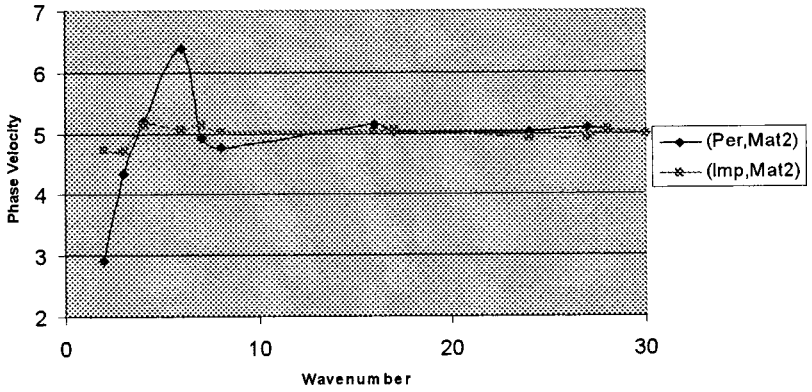


FIG. 6. Phase velocity as a function of wave number (symmetric motion, material 2).

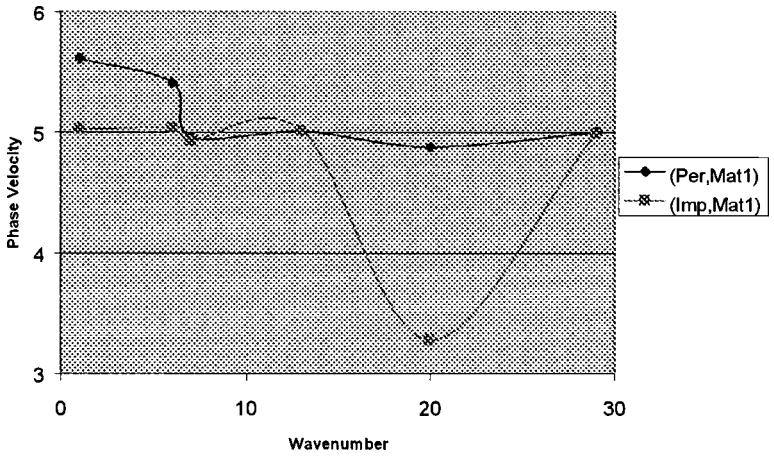


FIG. 7. Phase velocity as a function of wave number (antisymmetric motion, material 1).

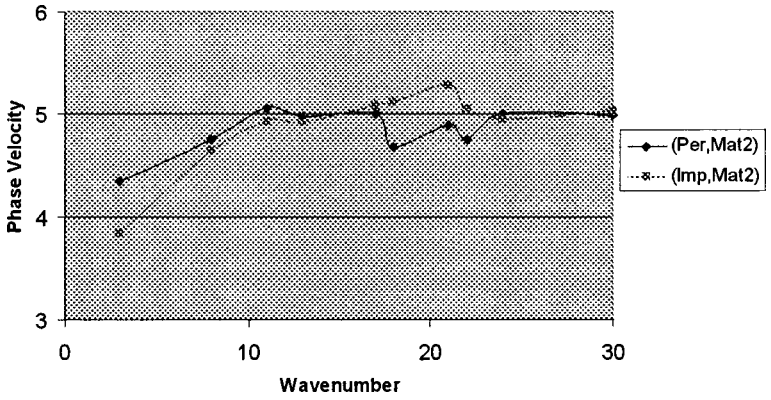


FIG. 8. Phase velocity as a function of wave number (antisymmetric motion, material 2).

Similarly, the frequency equation (51) for an impervious surface yields

$$|N_{ij}|=0 \quad (i,j=1,2,3), \tag{57}$$

where

$$\begin{aligned} &N_{11}=L_{11}, \quad N_{12}=L_{12}, \quad N_{13}=L_{13},\\ &N_{21}=L_{21}, \quad N_{22}=L_{22}, \quad N_{23}=L_{23},\\ &N_{31}=D_{31}D_{41}, \quad N_{32}=D_{32}D_{42}, \quad N_{33}=0. \end{aligned} \tag{58}$$

In (58), L_{ij} are defined in (56) and D_{ij} are defined in (38).

Equations (55) and (57) are analogous to the frequency equations of Rayleigh waves in a semi-infinite poroelastic solid for a pervious and an impervious surface, respectively.

A case of dissipative medium can also be considered (see Biot, 1956; Tajuddin and Sharma, 1980; Tajuddin, 1990) which needs further a massive amount of detailed analysis. We shall discuss such behavior later.

V. NUMERICAL RESULTS AND DISCUSSIONS

Two types of poroelastic materials are considered to carry out the computational work, one is sandstone saturated with kerosene, say material 1 (Fatt, 1959), the other is sandstone saturated with water, material 2 (Yew and Jogi, 1976), whose values are given by Table I.

Given material, each frequency equation is an implicit relation between nondimensional phase velocity and nondimensional wave number. Since this relation is transcendental

there exists uncountable number of roots (phase velocities) for each wave number. Phase velocity is computed for the values of wave number from 1 to 30 and presented graphically by Figs. 1–8. One can infer the following conclusions from the results:

(1) In case of pervious surface, material 1 values are, in general, less than that of material 2 for both symmetric and antisymmetric motions.
(2) In case of impervious surface, material 1 values are, in general, greater than that of material 2 for both symmetric and antisymmetric motions.
(3) If the values are analyzed using ''range'' a simple statistical measure of dispersion, the values of both pervious and impervious surfaces show opposite behavior with respect to type of material and type of motion.

ACKNOWLEDGMENTS

Thanks are due to Dr. Paul J. Remington, Associate Editor, for suggestions that improved the presentation of this paper. One of the authors (P.M.R.) wishes to acknowledge Council of Scientific and Industrial Research (CSIR), Government of India, for financial support in the form of Junior and Senior Research Fellowships.

Biot, M. A. (**1956**). ''The theory of propagation of elastic waves in fluid-saturated porous solid,'' J. Acoust. Soc. Am. **28**, 168–178.
Biot, M. A. (**1956**). ''General solutions of the equations of elasticity and consolidation for a porous material,'' J. Appl. Mech. **23**, 91–95.
Chein, N., and Hermann, G. (**1996**). ''Conservation Laws for Thermo or Poroelasticity,'' Trans. ASME, J. Appl. Mech. **63**, 331–336.
De Boer, R., and Ehlers, W. (**1988**). ''A historical review of formulation of porous media theories,'' Acta Mech. **79**, 1–8.
Eringen, A. C., and Suhubi, E. S. (**1975**). *Elastodynamics, Vol. 2, Linear Theory* (Academic, New York).
Ewing, W. M., Jardetzsky, W. S., and Press, F. (**1957**). *Elastic Waves in Layered Media* (McGraw-Hill, New York).
Fatt, I. (**1959**). ''The Biot–Willis elastic coefficients for a sandstone,'' J. Appl. Mech. **26**, 296–297.
Gardner, G. H. F. (**1962**). ''Extensional waves in fluid saturated porous cylinders,'' J. Acoust. Soc. Am. **34**, 36–40.
Jensen, O. E., Gluckberg, M. R., Sach, J. R., and Grotberg, J. B. (**1994**). ''Weakly non-linear deformation of thin poroelastic layer with a free surface,'' ASME J. Appl. Mech. **61**, 729–731.
Kolsky, H. (**1963**). ''Stress waves in solids,'' J. Sound Vib. **33**, 88–110.
Kumar, S. (**1960**). *Edge Waves in Plates, Proceedings of the International Symposium on Stress Wave Propagation in Materials* (Interscience, New York).
Miklowitz, J. (**1960**). ''Elastic wave propagation,'' Appl. Mech. Rev. **13**, 865–878.
Paria, G. (**1963**). ''Flow of fluid through porous deformable solids,'' Appl. Mech. Rev. **16**, 421–423.
Reddy, P. M., and Tajuddin, M. (**1996**). *Proceedings of the Third International Conference on Vibration Problems* (University of North Bengal, India), pp. 92–94.
Reddy, P. M., and Tajuddin, M. (**2000**). ''Exact analysis of the plane strain vibrations of thick walled hollow poroelastic cylinders,'' Int. J. Solids Struct. **37**, 3439–3456.
Taber, L. A. (**1992**). ''A theory for transverse deflection of poroelastic plates,'' J. Appl. Mech. **59**, 628–634.
Tajuddin, M. (**1978**). ''Vibrations in fluid—saturated porous elastic cylinders,'' Rev. Roum. Sci. Tech., Ser. Mec. Appl. **23**, 371–379.
Tajuddin, M. (**1982**). ''Rayleigh waves on a concave cylindrical poroelastic surface,'' Indian J. Pure Appl. Math. **13**, 1278–1282.
Tajuddin, M. (**1984**). ''Rayleigh waves in a poroelastic half-space,'' J. Acoust. Soc. Am. **75**, 682–684.
Tajuddin, M. (**1990**). ''On torsional impulsive loading of poroelastic surface,'' Act. Geophys. Polon. **38**, 159–167.
Tajuddin, M., and Ahmed, S. I. (**1991**). ''Dynamic interaction of a poroelastic layer and half-space,'' J. Acoust. Soc. Am. **89**, 1169–1175.
Tajuddin, M., and Moiz, A. A. (**1984**). ''Rayleigh waves on a convex cylindrical poroelastic surface,'' J. Acoust. Soc. Am. **76**, 1252–1254.
Tajuddin, M., and Sarma, K. S. (**1980**). ''Torsional vibrations of poroelastic cylinders,'' ASME J. Appl. Mech. **47**, 214–216.
Thomson, M., and Willis, J. R. (**1991**). ''A reformation of the equations of anisotropic poroelasticity,'' ASME J. Appl. Mech. **58**, 612–616.
Yew, C. H., and Jogi, P. N. (**1976**). ''Study of wave motions in fluid-saturated porous rocks,'' J. Acoust. Soc. Am. **60**, 2–8.

An approximate Green's function for a locally excited fluid-loaded thin elastic plate

Daniel T. DiPerna[a] and David Feit[b]
Carderock Division NSWC, West Bethesda, Maryland 20817

(Received 14 November 2002; revised 4 April 2003; accepted 7 April 2003)

In the classic treatment of the line-driven, fluid-loaded, thin elastic plate, a branch cut integral typically needs to be evaluated. This branch cut arises due to a square root operator in the spectral form of the acoustic impedance. In a previous paper [J. Acoust. Soc. Am. **110**, 3018 (2001)], DiPerna and Feit developed a methodology, complex layer analysis (CLA), to approximate this impedance. The resulting approximation was in the form of a rational function, although this was not explicitly stated. In this paper, a rational function approximation (RFA) to the acoustic impedance is derived. The advantage of the RFA as compared to the CLA approach is that a smaller number of terms are required. The accuracy of the RFA is examined both in the Fourier transform domain and the spatial domain. The RFA is then used to obtain a differential relationship between the pressure and velocity on the surface of the plate. Finally, using the RFA in conjunction with the equation of motion of the plate, an approximate expression for the Green's function for a line-driven plate is obtained in terms of a sum of propagating and evanescent waves. Comparisons of these results with the numerical inversion of the exact integral show reasonable agreement. [DOI: 10.1121/1.1577549]

PACS numbers: 43.40.Rj, 43.40.Jx [EGW]

I. INTRODUCTION

One of the fundamental problems of structural acoustics is that of a fluid-loaded thin elastic plate, excited locally by time-harmonic forces. A numerically efficient description of the wave mechanisms that exist in the plate response and radiation is useful in determining the structural acoustic behavior of more realistic structures. This problem has therefore been the subject of many studies over the last few decades.[1–4] An excellent review of the subject is included in David Crighton's Institute of Acoustics Rayleigh Lecture.[5]

The equation of motion for the line-driven fluid-loaded plate is simply the equation of motion for the *in-vacuo* plate plus a term corresponding to the reaction force exerted on the plate by the fluid. The fluid-loading term is the solution of the Helmholtz equation, which satisfies the Sommerfeld radiation condition as well as continuity of normal velocity. This can be written as an integral involving the Green's function of the wave equation and the velocity of the plate.[6] The equation of motion for the plate subject to fluid loading thus becomes an integro-differential equation for the plate velocity distribution.

The classical treatment is to Fourier transform the integro-differential equation, solve for the velocity transform, and then invert the velocity transform, which is equal to the transform of the applied force divided by the sum of the structural and acoustic spectral impedances. The latter, hereafter referred to as the acoustic impedance, is equal to the ratio of the transform of the surface pressure to that of the velocity and contains a square root operator. This operator introduces a multivalued function of the transform parameter, while the structural spectral impedance is a polynomial function of the transform parameter. Because the acoustic impedance contains branch points, an inverse transform involves a branch cut integral which, must be either numerically or asymptotically evaluated.[7]

In a previous paper,[8] we introduced a novel approach, referred to as complex layer analysis (CLA) to help overcome this complication. CLA replaces the semi-infinite fluid medium by a finite thickness layer of fluid, resulting in a spectral form of the acoustic impedance, which is no longer multivalued. The spatial domain solution for the line-driven plate/fluid system can then be obtained as an infinite modal sum over a discrete spectrum. These correspond to the modes of the fluid layer interacting with the elastic plate. In Ref. 8, it was shown that by assuming the layer thickness to be complex, the required modal sum is reduced to a smaller number of terms. The utility of the solution so obtained is a gain in efficiency of calculating the response field to a localized force, e.g., a point or line-drive excitation. This Green's function facilitates the solution of problems for more realistic configurations; e.g., fluid-loaded plates complicated by the presence of attached substructures such as stiffeners, or changes in the plate's material properties, etc.

The representation of the Green's function in terms of a finite summation of pole contributions obtained using the CLA analysis was possible because the approximate acoustic impedance used there was expressed as a rational function, although this was not explicitly stated there. Motivated by this understanding, in this study, we proceeded using a slightly different approach. Instead of using CLA whose final result shows that the acoustic impedance can be approximated by a rational function, we numerically find a rational function approximation (RFA) to the acoustic impedance. This is accomplished by approximating the acoustic imped-

[a]Electronic mail: dipernadt@nswccd.navy.mil
[b]Electronic mail: feitd@nswccd.navy.mil

ance as a ratio of polynomials.[9] By using this form of the acoustic impedance, the spectral form of the velocity Green's function is no longer multivalued, and has simple poles in the complex transform plane. However, in this case the number of poles is limited and dependent on the order of the polynomials used in the approximation.

Effectively, what is accomplished by the RFA approach is a replacement of the branch-cut singularity in the exact acoustic impedance by a series of zeros and poles in the approximated acoustic impedance. In mathematical terms the continuous spectrum operator is replaced (approximated) by a discrete spectrum operator. In general the discrete part of the spectrum produces surface waves, while the continuous part produces cylindrical or space waves, the latter depending on the number of spatial dimensions considered.[10]

This study, after first introducing the RFA for the acoustic impedance, derives a new differential relationship between surface pressure and plate normal velocity. The RFA version of the acoustic impedance is then used to evaluate the surface pressure Green's function, that is the surface pressure due to a line distribution of velocity on an otherwise rigid surface.

Using this new formulation, we then compare the results of varying order RFA's on the dispersion curves for the fluid-loaded plate, identify the various wave mechanisms, and compare the Green's functions for the line-driven fluid-loaded elastic plate using the RFA approach with those calculated by numerical integration.

II. FORMULATION OF THE PROBLEM

Using similar notation [except $u(x)$, the plate velocity replaced by $v(x)$] and proceeding in the same manner as Ref. 8, we rewrite Eq. (3) of that reference, in the following form:

$$\tilde{v}(k_x)=\frac{\tilde{f}(k_x)}{\tilde{Z}_p(k_x)+\tilde{Z}_a(k_x)}, \tag{1}$$

where the plate and acoustic impedances appearing in the denominator of (1) are given by

$$\tilde{Z}_p(k_x)=-i\omega m\left(1-\frac{k_x^4}{k_f^4}\right),$$
$$\tilde{Z}_a(k_x)=\frac{\rho\omega}{\sqrt{k_0^2-k_x^2}}=\rho c\,\frac{1}{\sqrt{1-(k_x/k_0)^2}}, \tag{2}$$

m is the mass per unit area of the plate material, ρ is the fluid medium density, c is the speed of sound in the fluid, ω is the excitation frequency, $k_f=(m\omega^2/D)^{1/4}$ is the *in-vacuo* plate flexural wave number, D is the flexural rigidity of the plate, $k_0=\omega/c$ is the acoustic wave number, k_x is the transform parameter and $\tilde{f}(k_x)$ is the Fourier transform of the excitation function. The Fourier transform pair is given by

$$\tilde{y}(k_x)=\int_{-\infty}^{\infty} y(x)\exp(-ik_x x)\,dx,$$
$$y(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{y}(k_x)\exp(ik_x x)\,dk_x.$$

When, in a later section, we evaluate the Green's function, i.e., the response to a unit force line-excitation at the origin $f(x)=\delta(x)$, the transform of this becomes $\tilde{f}(k_x)=1$.

Both of the spectral impedance functions appearing in (2) are even functions of the transform parameter k_x, but as stated previously, $\tilde{Z}_a(k_x)$ is a multivalued function having branch points at $k_x=\pm k_0$. This complicates the inversion to the spatial domain, and precludes the expression of the Green's function in a closed form; it is primarily this complication that the current approximation procedure addresses.

III. RATIONAL FUNCTION APPROXIMATION OF THE ACOUSTIC IMPEDANCE

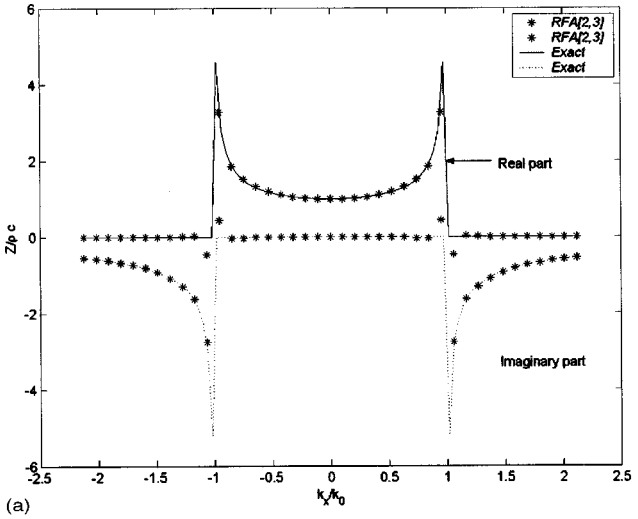
At this point we introduce the RFA of the acoustic impedance by writing the impedance as a ratio of polynomials, given by the formula

$$\tilde{Z}_a(k_x)\approx\tilde{Z}_a^{[L,J]}(k_x)=\frac{\rho c\sum_{m=0}^{L} a_{2m}(k_x/k_0)^{2m}}{1+\sum_{n=1}^{J} b_{2n}(k_x/k_0)^{2n}}=\rho c\,\frac{N^{[L]}(k_x/k_0)}{D^{[J]}(k_x/k_0)}, \tag{3}$$

where a_{2m} and b_{2n} are determined, as detailed in the Appendix, by minimizing the least square error difference between the actual function and its approximation over a finite interval of the real axis. The superscripts appearing in the RFA refer to the polynomial powers in the numerator and denominator, respectively. For the remainder of this study these approximations are labeled as RFA$[L,J]$. As an example, for the lowest order approximation RFA[2,3] we take $L=2$, $J=3$ in Eq. (3) making the numerator and denominator fourth- and sixth-order polynomials, respectively. The higher order approximation is denoted as RFA[3,4] with sixth- and eighth-order polynomials in the numerator and denominator, respectively.

The specific requirements on the RFA, i.e., the order of the polynomials and the approximation interval for any particular application are discussed in the Appendix. In Fig. 1 we show two different approximations (RFA[2,3] and RFA[3,4]) defined by Eq. (3) evaluated along the real axis compared with the exact impedance function.

Using the definition of the acoustic impedance as the ratio of the transform of the pressure to that of the velocity, and using the RFA[2,3] for illustrative purposes we find that



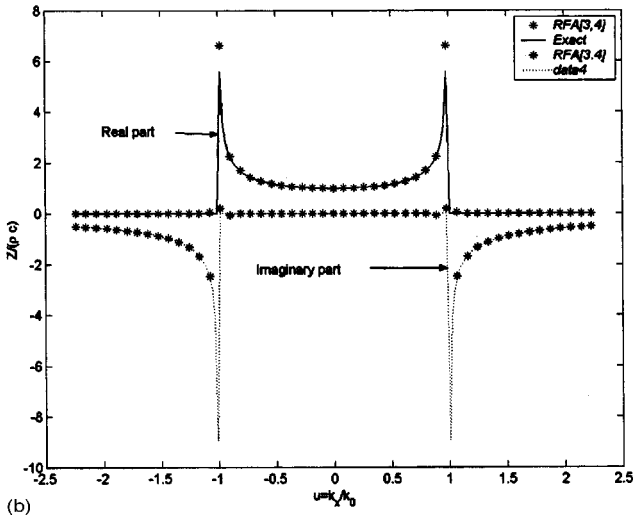


FIG. 1. (a) Real and imaginary parts of normalized spectral acoustic impedance calculated for RFA[2,3] compared to results calculated from exact expression. (b) Real and imaginary parts of normalized spectral acoustic impedance calculated for RFA[3,4] compared to results calculated from exact expression.

$$\left(1+b_2\left(\frac{k_x}{k_0}\right)^2+b_4\left(\frac{k_x}{k_0}\right)^4+b_6\left(\frac{k_x}{k_0}\right)^6\right)\tilde{P}(k_x) = \rho c\left(a_0+a_2\left(\frac{k_x}{k_0}\right)^2+a_4\left(\frac{k_x}{k_0}\right)^4\right)\tilde{V}(k_x). \tag{4}$$

Taking the inverse transform of the above equation yields a newly derived differential pressure/velocity relationship in the spatial domain of the form

$$-\left(\frac{b_6}{k_0^6}\right)\frac{d^6p}{dx^6}+\left(\frac{b_4}{k_0^4}\right)\frac{d^4p}{dx^4}-\left(\frac{b_2}{k_0^2}\right)\frac{d^2p}{dx^2}+p = \rho c\left[\left(\frac{a_4}{k_0^4}\right)\frac{d^4v}{dx^4}-\left(\frac{a_2}{k_0^2}\right)\frac{d^2v}{dx^2}+a_0v\right]. \tag{5}$$

Note that as frequency increases the coefficients multiplying the higher order derivatives become smaller suggesting that the approximations introduced here become more valid with increasing frequency.

Equation (5) combined with the equation of motion of the plate will yield a coupled set of differential equations for the surface pressure and normal velocity. This contrasts with the classic approach, discussed earlier, which yielded an integro-differential equation governing the relationship between surface pressure and normal velocity.

The fact that the RFA has introduced higher order spatial derivatives implies that there must be corresponding higher order boundary conditions. For the purposes of this paper, that of obtaining approximate Green's function for the fluid-loaded plate, the boundary conditions are simply radiation conditions, i.e., that all waves are outgoing from the location of the point drive. This always gives the appropriate number of conditions no matter what the order of the approximation. These boundary conditions are implicit in the contour choice in the Green's function evaluation in the following sections.

This newly derived formulation of the classical structural acoustic problem consists of two coupled ordinary differential equations on the surface pressure and the plate velocity. While not pursued here, these coupled ordinary differential equations can be cast into a state space formulation. Here dependent variables (states), i.e., the surface pressure, velocity and their derivatives are arranged in a column vector called the ''state-vector.'' This vector defines the complete mechanical and fluid states at each section. The previously described differential equations are then used to describe the transmission of states from one location to another. This state space formulation is very well suited for handling structures made inhomogeneous by the presence of attached substructures.

IV. PRESSURE PER UNIT VELOCITY GREEN'S FUNCTION

One way of judging the relative efficacy of the RFA approach is to evaluate the inverse transform of the acoustic impedance in Eq. (2) not using the RFA approach. The inverse transform can be integrated analytically yielding the result[11]

$$\frac{p(k_0x)}{\rho c}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\exp(iuk_0x)}{\sqrt{1-u^2}}\,du=\frac{iH_0(k_0x)}{2}. \tag{6}$$

The function obtained in Eq. (6) can be interpreted as the surface pressure distribution generated by a unit velocity concentrated at the origin $x=0$, i.e., the surface pressure Green's function.

Now, by introducing the RFA approximations for the acoustic impedance and spatially inverting the RFA approximation as given by Eq. (3), we can obtain a function, which approximates this Green's function. As an example, for the RFA[2,3] case the denominator of the integrand is a sixth-order polynomial in $u=(k_x/k_0)$ giving rise to six poles, three of which are in the first quadrant of the complex u

plane, and three in the third quadrant. Using the theory of residues the inverse transform can be written as

$$\frac{\breve{p}(k_0x)}{\rho c} \approx \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{N^{[L]}(u)}{D^{[J]}(u)} \exp(iuk_0x)du$$

$$= i\sum_{n=1}^{J} R(u_n)\exp(iu_n|x-x_0|), \tag{7}$$

where $R(u_n)$ is the residue calculated at the poles (u_n) of the integrand in the first quadrant, i.e., those with $\Im(u_n)\geqslant 0$.

The results of Eq. (7) compared to the exact function obtained using Eq. (6) is shown in Fig. 2 for two different approximation orders. Both approximations are valid for higher values of the argument. The error at the lower values of the argument is governed by the maximum wavenumber used in the approximation process (see Appendix.)

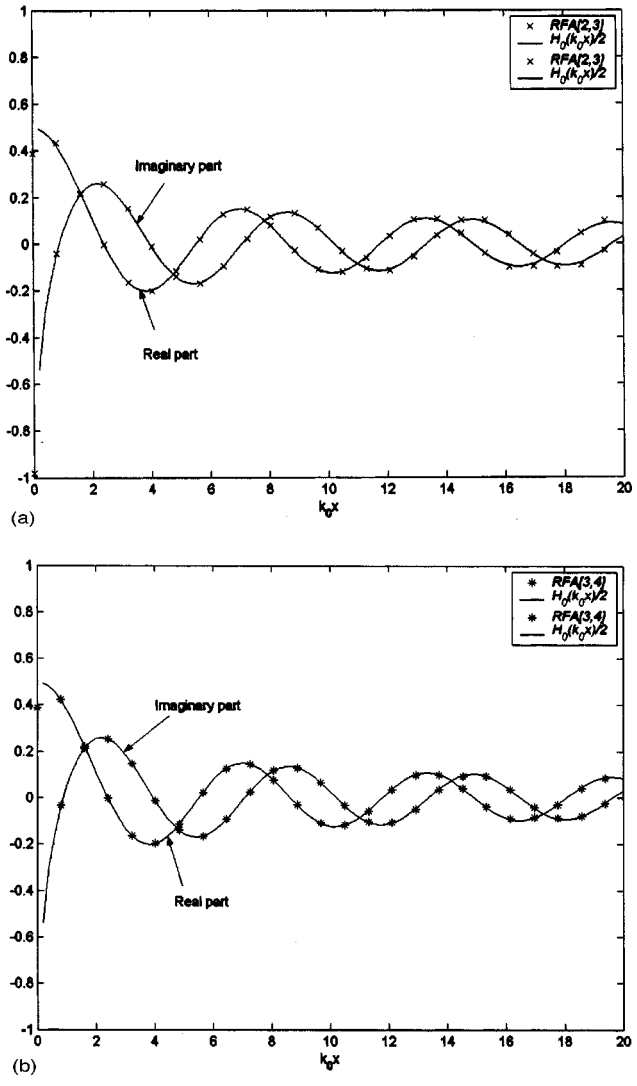


FIG. 2. (a) Real and imaginary parts of surface pressure Green's function calculated for RFA[2,3] compared to results calculated from exact expression $H_0(k_0x)/2$. (b) Real and imaginary parts of surface pressure Green's function calculated for RFA[3,4] compared to results calculated from exact expression $H_0(k_0x)/2$.

V. GREEN'S FUNCTION FOR THE FLUID-LOADED PLATE VELOCITY USING THE RFA APPROACH

Taking the applied loading as $\delta(x-x_0)$, and introducing the RFA into Eq. (1), the plate velocity Green's function is obtained in the form

$$\breve{G}(x,x_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{D^{[J]}(k_x/k_0)\exp[ik_x(x-x_0)]}{Z_p(k_x)D^{[J]}(k_x/k_0)+N^{[L]}(k_x/k_0)}dk_x. \tag{8}$$

The approximate Green's function, estimated using the RFA, is written as $\breve{G}(x,x_0)$ to distinguish it from the exact Green's function obtained by numerical inversion of the exact velocity transform. The denominator of the integrand is now a polynomial function of the transform parameter, so that the inverse transform is readily evaluated using residue theory. As stated earlier, for illustrative purposes we use the RFA[2,3]. In this case the denominator is a tenth-order polynomial (fourth-order from the structural impedance and sixth-order coming from the acoustic impedance approximation).

Again, introducing the nondimensional variable $u = k_x/k_0$, the dispersion relationship for RFA[2,3] becomes

$$u^{10}+\alpha_8u^8+\alpha_6u^6+\alpha_4u^4+\alpha_2u^2+\alpha_0=0, \tag{9a}$$

where

$$\alpha_8=\frac{b_4}{b_6}, \quad \alpha_6=\left(\frac{b_2}{b_6}-\frac{1}{M^4}\right), \quad \alpha_4=\left(\frac{1}{b_6}-\frac{b_4}{M^4b_6}-\frac{i\epsilon a_4}{M^6b_6}\right),$$
$$\alpha_2=-\left(\frac{b_2}{M^4b_6}+\frac{i\epsilon a_2}{M^6b_6}\right), \quad \alpha_0=\left(\frac{i\epsilon a_0}{M^6b_6}+\frac{1}{M^4b_6}\right). \tag{9b}$$

In Eq. (9b) we have introduced the notation: $\epsilon = \rho c/(\omega_c m)$, $M=\sqrt{\omega/\omega_c}$, where $\omega_c=\sqrt{12c^2}/(hc_p)$ is the coincidence frequency. ϵ is known as the ''intrinsic fluid loading parameter'' equal to 0.129 for a steel plate in water, while M is in the form of a Mach number, i.e., the ratio of the frequency dependent flexural wave speed to the acoustic wave speed of water. While Eq. (9a) must be solved numerically as a function of frequency, the solution of a polynomial equation is algorithmic. The solutions are complex and denoted as u_j, where the real part is inversely proportional to the wave speed while the imaginary part is proportional to the spatial attenuation factor. These solutions are used to evaluate the velocity Green's function by the theory of residues, which takes on the form

$$\breve{G}(x,x_0) = i\sum_{j=1}^{J+2} R(u_j)\exp(iu_jk_0|x-x_0|), \tag{10}$$

where $R(u_j)$ is the residue calculated at the roots of Eq. (9a) satisfying the condition $\Im(u_j)\geqslant 0$. The Green's function given by Eq. (10) is now in the form of a finite set of propagating surface and exponentially decaying disturbances launched at the excitation point.

Figure 3(a) is a gray-scale plot of the absolute value of the plate velocity transform using the exact expressions as

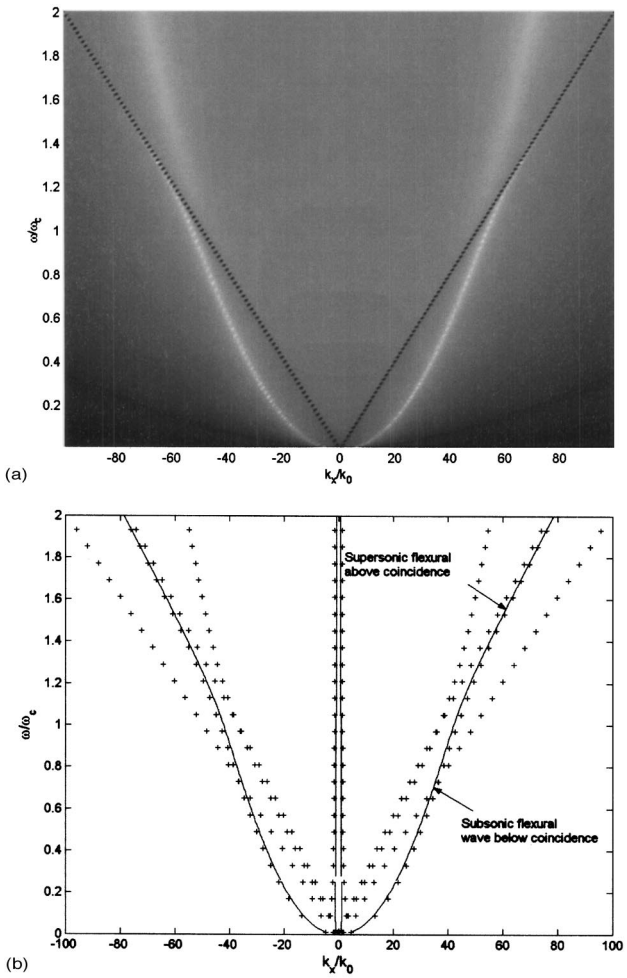


FIG. 3. (a) Gray-scale map of absolute value of velocity transform using exact expression Eq. (1). (b) Dispersion curves using RFA[2,3] (plotting symbol "+"), real parts of the roots of Eq. (9a) (plotted as "solid line").

given by Eqs. (1) and (2) evaluated at real values of wave number k_x. The light lines are large values corresponding to the poles associated with the fluid-loaded flexural waves. The dark lines correspond to zeros associated with the high impedance of the fluid at the acoustic wave number. In Fig. 3(b) we compare the dispersion curves calculated at discrete frequencies using the exact theory with those calculated from the approximate RFA theory. For the exact theory there are four dispersion curves (two each for the disturbances traveling in the positive and negative directions), while for the RFA[2,3] approximation, there are 10 curves. The additional dispersion curves arise from the three poles that are used to approximate the branch cut integral of the classical approach. These do not manifest themselves in Fig. 3(a) because they have large imaginary components. One of the curves is labeled as the fluid-loaded flexural wave, which is seen to merge with the acoustic line in the vicinity of the coincidence frequency $\omega/\omega_c=1$. Above coincidence there is another disturbance that follows the flexural wave curve, and is supersonic. This wave appears to emanate from a disturbance that below coincidence was close to the acoustic wave, but slightly subsonic. The exact and approximated dispersion curves are in very good agreement. These wave mechanisms have been previously well documented in the literature.[6]

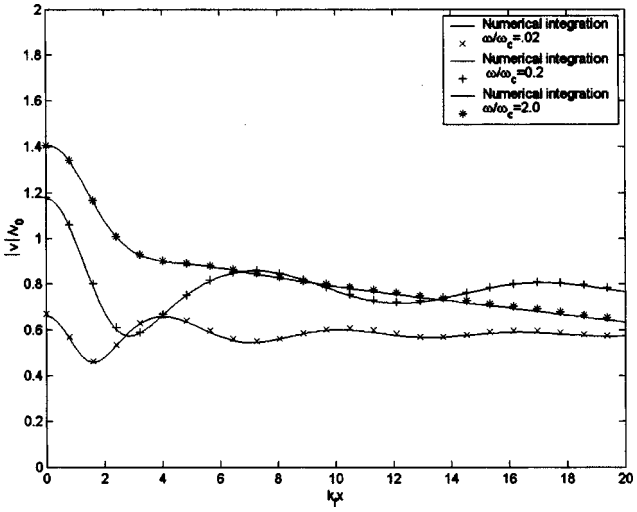


FIG. 4. Plate velocity normalized by $v_0=k_f/(4\omega m)$ as a function of nondimensional distance $k_f x$ for three different frequencies calculated using RFA[2,3] compared to inverse of velocity transform calculated numerically.

Figure 4 compares the results of using the RFA[2,3] in calculating the normalized plate velocity as a function of normalized distance from the drive point to that obtained using a numerical inversion of the plate velocity Fourier transform using the exact expression for the acoustic impedance. Results are shown for three frequencies (two below coincidence and one above coincidence). The solid lines are from the numerical integration while the marked lines are obtained using the RFA[2,3]. The agreement is excellent for the three frequencies considered.

VI. SUMMARY

A new approach is used to efficiently calculate the Green's function for a fluid-loaded elastic plate. This is accomplished by introducing a rational function approximation of the acoustic impedance. Calculation of the velocity response using this approximation is numerically very efficient and is in excellent agreement with earlier calculations using numerical integration to invert the transform solution. The same approximation is also used to derive a new differential surface pressure/velocity relationship which we anticipate will be of great use in developing a state-space formulation for more complex structural acoustic configurations.

ACKNOWLEDGMENT

This work was sponsored by the Carderock Division Intra-Laboratory Independent Research (ILIR) Program, administered by Dr. John Barkyoumb.

APPENDIX

This appendix will briefly discuss how the acoustic impedance operator is approximated by a rational function. The acoustic impedance is given by

$$\tilde{Z}_a(k_x)=\rho c s(u),$$

$$s(u)=\frac{1}{\sqrt{1-u^2}}, \tag{A1}$$

$u = k_x / k_0$.

As stated in the text, we want to approximate $s(u)$ as follows:

$$s(u) \approx s^{[L,J]}(u) = \frac{\sum_{m=0}^{L} a_{2m} u^{2m}}{1 + \sum_{n=1}^{J} b_{2n} u^{2n}} \tag{A2}$$

which can be rewritten as

$$\sum_{m=0}^{L} a_{2m} u^{2m} - s(u) \sum_{n=1}^{J} b_{2n} u^{2n} = s(u). \tag{A3}$$

After evaluating this expression at a number of points in the range of u (i.e., $-u_{\max} < u < u_{\max}$) the above equation becomes a linear system of equations for the coefficients a_{2m} and b_{2n}. By choosing the number of points to be much larger than the number of unknowns, this can be solved in the least-squares sense.

The range of u and the number of unknowns depend on the specific problem at hand. In general, the approximation breaks down beyond the range over which Eq. (A3) was minimized. This is because for large arguments, $s(u)$ varies as $1/u$ whereas $s^{[L,J]}(u)$ goes as $1/u^2$ if we choose $J = L + 1$. In order to calculate the total impedance of a fluid/structure interaction problem $u_{\max} \approx 1.25 k_f / k_0$, where k_f is the maximum flexural wave number over the entire frequency range, which can be approximated by $k_f / k_0 \approx 1/\sqrt{M}(1 + \epsilon/4M^2)$. For the approximations used in this paper, we used the following sets of coefficients.

Coefficients for RFA[2,3] with $N = 201$ points covering the range of u.

$a_0 = 0.982\,668\,48 + 0.015\,620\,86\,i$
$a_2 = -0.879\,689\,45 + 0.633\,845\,67\,i$
$a_4 = -0.163\,038\,52 - 0.388\,082\,90\,i$
$b_2 = -1.540\,604\,70 + 0.721\,624\,99\,i$
$b_4 = 0.445\,236\,25 - 0.701\,980\,96\,i$
$b_6 = 0.068\,849\,33 - 0.008\,484\,85\,i$

Coefficients for RFA[3,4] with $N = 201$ points covering the range of u.

$a_0 = 1.004\,407\,70 - 0.001\,006\,10\,i$
$a_2 = -1.459\,966\,56 + 0.944\,011\,64\,i$
$a_4 = 0.184\,554\,13 - 0.897\,248\,21\,i$
$a_6 = 0.180\,746\,46 - 0.004\,164\,37\,i$
$b_2 = -1.895\,033\,23 + 0.945\,320\,63\,i$
$b_4 = 0.634\,904\,74 - 1.451\,518\,29\,i$
$b_6 = 0.258\,335\,31 + 0.480\,932\,81\,i$
$b_8 = -0.005\,177\,51 + 0.021\,232\,37\,i$

[1] L. Y. Gutin, "Sound radiation from an infinite plate excited by a normal point force," Sov. Phys. Acoust. **10**, 369–371 (1965).
[2] G. Maidanik and E. Kerwin, "Influence of fluid loading on the radiation from infinite plates below the critical frequency," J. Acoust. Soc. Am. **40**, 1036–1038 (1967).
[3] D. Feit, "Pressure radiated by a point excited plate," J. Acoust. Soc. Am. **40**, 1489–1494 (1967).
[4] D. Crighton, "Force and moment admittance of plates under arbitrary loading," J. Sound Vib. **20**, 209–218 (1972).
[5] D. Crighton, "Fluid loading: the interaction between sound and vibration," Proceedings of the Institute of Acoustics **10**, Part 2, 1–35 (1988).
[6] M. C. Junger and D. Feit, *Sound, Structures, and their Interaction*, 2nd ed. (Acoustical Society of America, 1993), Chap. 8, pp. 235–278.
[7] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, New Jersey, 1973), pp. 455–476.
[8] D. DiPerna and D. Feit, "An approximate analytic solution for the radiation from a line-driven fluid-loaded plate," J. Acoust. Soc. Am. **110**, 3018–3024 (2001).
[9] *Numerical Recipes in C; The Art of Scientific Computing*, edited by W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (Cambridge University Press, Cambridge, MA, 1992).
[10] B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956), Chap. 2, pp. 57–133.
[11] L. S. Gradshtyn and I. M. Rhyzhik, *Table of Integrals, Series, and Products*, 5th ed. (Academic, New York, 1980), 8.411 and 8.415, pp. 961–965.

Transmission loss predictions for dissipative silencers of arbitrary cross section in the presence of mean flow

Ray Kirby[a)]
Department of Mechanical Engineering, Brunel University, Uxbridge, Middlesex UB8 1PH, United Kingdom

(Received 20 September 2001; revised 20 February 2003; accepted 25 April 2003)

A numerical technique is developed for the analysis of dissipative silencers of arbitrary, but axially uniform, cross section. Mean gas flow is included in a central airway that is separated from a bulk reacting porous material by a concentric perforate screen. The analysis begins by employing the finite element method to extract the eigenvalues and associated eigenvectors for a silencer of infinite length. Point collocation is then used to match the expanded acoustic pressure and velocity fields in the silencer chamber to those in the inlet and outlet pipes. Transmission loss predictions are compared with experimental measurements taken for two automotive dissipative silencers with elliptical cross sections. Good agreement between prediction and experiment is observed both without mean flow and for a mean flow Mach number of 0.15. It is demonstrated also that the technique presented offers a considerable reduction in the computational expenditure when compared to a three-dimensional finite element analysis. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1582448]

PACS numbers: 43.50.Gf, 43.20.Mv, 43.20.Hq [MRS]

I. INTRODUCTION

Dissipative silencers are effective for attenuating broadband noise and are commonly deployed in automotive exhaust and HVAC systems. A dissipative silencer often takes on a complex geometrical shape, for example, in an automotive exhaust system elliptic cross sections are common. Modeling complex silencer geometries presents a considerable challenge, especially if one assumes the porous material to be bulk reacting. Inevitably, numerical techniques have found favor for modeling irregular geometries and for dissipative silencers the finite element method (FEM) is used widely. A general application of the FEM to dissipative silencer design was presented by Peat and Rathi,[1] who reported transmission loss predictions for two axisymmetric exhaust silencers, both with and without mean flow in the central airway. The analysis of Peat and Rathi is capable of modeling fully arbitrary silencer geometries, although, in general, this requires the use of a three-dimensional finite element mesh.[2] Most silencer geometries are not, however, always of a fully arbitrary shape; in fact, most dissipative silencers usually contain an axially uniform cross section. For such a silencer it is desirable to take advantage of the uniform geometry and to avoid the significant CPU expenditure associated with a three-dimensional finite element model. One possible solution is to apply the so-called point collocation technique suggested by Astley *et al.*[3] This method is versatile enough to cope with an arbitrary cross section but also promises to economize on the CPU expenditure when compared to the method of Peat and Rathi.[1]

The model reported here examines a ''straightthrough'' dissipative silencer containing an axially uniform, but arbitrarily shaped, cross section. The model includes mean flow in the central airway and also a perforated screen, separating the porous material from the central airway, since this has been shown also to influence the silencer performance.[4] A uniform silencer facilitates the reduction of the problem from three to two dimensions and in the process potentially reduces the CPU expenditure. Thus, the silencer chamber studied here is assumed first to be infinite in length and an eigenvalue analysis is performed. Subsequently the silencer transmission loss is computed by matching the expanded acoustic pressure and velocity fields at the entry/exit planes of the silencer chamber.

The relative simplicity of an eigenvalue analysis, particularly when compared to a three-dimensional approach, has meant that computing modal attenuation rates for dissipative silencers has proved popular, although very few studies progress to calculating silencer transmission loss. For example, Astley and Cummings[5] use the FEM to compute modal attenuation rates in dissipative silencers of a rectangular cross section, adding the effects of mean flow in the central airway. The method of Astley and Cummings was later applied to automotive silencer design by Rathi,[6] who obtained modal attenuation rates for silencers of an elliptic cross section. Both studies do, however, omit the effects of a perforate and, more importantly, neither progress to predicting the silencer transmission loss. A number of alternative numerical eigenvalue formulations have also been sought for elliptical cross sections, examples include the Rayleigh Ritz approach of Cummings[7] and the point matching technique of Glav.[8] These alternative formulations do, however, compromise, to some extent, the versatility and robustness of the FEM; the analysis of Cummings is restricted to the fundamental mode only, the method of Glav is very sensitive to silencer geometry and the collocation grid chosen. Moreover, Glav omits both mean flow and a perforate while Cummings

[a)]Electronic mail: ray.kirby@brunel.ac.uk

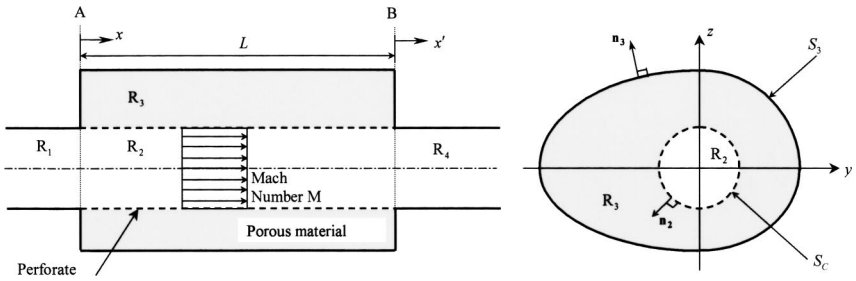


FIG. 1. Geometry of silencer.

omits a perforate, and neither study progresses to computing silencer transmission loss.

To predict silencer transmission loss, after first performing an eigenvalue analysis one begins by expressing the acoustic pressure and velocity fields on either side of a discontinuity (the inlet and outlet planes in a uniform straight through the silencer) as a modal expansion in which only the modal amplitudes are unknown. These are then determined by matching acoustic pressure and normal velocity across each discontinuity. This approach is commonly known as mode matching and has been applied successfully to duct acoustics problems, for example, Åbom[9] implemented an analytic mode matching technique when modeling a reactive exhaust silencer. In general, the method does, however, depend upon finding a transfer matrix **T** for the silencer, whose elements t_{ij} decay rapidly with increasing i, j. Without this property the solution to the truncated system of matching conditions may bear little resemblance to the solution of the physical problem. The decay of elements t_{ij} largely depends upon the weighting function chosen for the matching scheme. Åbom[9] studied a problem in which the underlying eigen-sub-system is Sturm–Liouville,[9] thus, if one chooses the modal eigenfunctions as weighting functions, modal orthogonality guarantees the rapid decay of elements t_{ij}. The underlying eigen-sub-system for a dissipative silencer is, however, non-Sturm–Liouville and choosing the modal eigenfunctions as weighting functions does not necessarily guarantee a convergent system of equations for a general class of problem. This problem may be addressed by substituting a suitable orthogonality relation that, in effect, restores modal orthogonality. Such an approach was adopted by Glav,[10] who successfully used an orthogonality relation to apply mode matching to a dissipative silencer of an arbitrary cross section. To arrive at a transfer matrix, Glav utilizes an appropriate orthogonality relation that, crucially, is valid only for zero mean flow. To extend the approach of Glav to include mean flow would require the solution of a system of equations in which the chosen weighting function is not orthogonal. Of course, the system of equations would remain tractable, however, convergence to a solution characteristic of the physical problem may not necessarily be achieved. This behavior is apparent in the results of Cummings and Chang[11] in which the good agreement between prediction and experiment is observed, but only under certain conditions; in this case at higher frequencies. At lower frequencies, when mean flow is present, predictions do not tend toward zero transmission loss, as one would expect. It is possible that this behavior is caused by the absence of an orthogonality relation in the analysis of Cummings and Chang[11] when mean flow is present, and so the subsequent solution bears little resemblance to the physical problem. In the absence of a suitable orthogonality relation for the present class of problem, caution is therefore exercised and an alternative method is investigated.

A straightforward alternative to analytic mode matching is to use a numerical matching technique. The method of point collocation, implemented by Astley *et al.*[3] in the study of air-conditioning ducts, appears well suited to the automotive silencer design. The technique involves equating velocity and pressure fields at discrete points, or nodal locations, on the cross section of the silencer, rather than integrating over the whole section, as is the case when matching analytically. Naturally, matching numerically cannot be expected to be as accurate as matching analytically, however, with a suitable choice of collocation points Kirby and Lawrie[12] demonstrated that, for a rectangular duct lined on opposite walls, excellent agreement with analytic mode matching predictions is possible. Although the study of Kirby and Lawrie omitted mean flow, their results do appear to vindicate the application of point collocation to the current problem. Of course, when implementing point collocation it is convenient first to perform an eigenvalue analysis using the FEM. The collocation points, over which pressure and velocity are matched, may then be chosen at any location over the transverse cross section, although the number of collocation points must not exceed the number of nodes in the original FE mesh. Thus, the analysis presented here first implements a finite element eigenvalue analysis, based on the method of Astley and Cummings[5] (with the addition of a perforated screen), and then implements a numerical point collocation matching scheme. Predictions are compared with experimental measurements taken for dissipative exhaust silencers with elliptical cross sections.

II. GOVERNING EQUATIONS

The dissipative silencer consists of a concentric perforated tube surrounded by an (isotropic) porous material of arbitrary cross section (see Fig. 1). The silencer chamber, which has a length L, is assumed to be uniform along its length, the outer walls of which are assumed to be rigid and impervious. The inlet and outlet pipes (regions R_1 and R_4) are identical, each having a circular cross section with rigid, impervious, walls.

Prior to matching acoustic pressure and velocity at each axial discontinuity, an eigenvalue analysis is required, both

for the silencer chamber and for the inlet/outlet pipes. Finding the eigenvalues and associated eigenvectors for the inlet/outlet pipes is straightforward so listed below is the eigenvalue analysis for the chamber only.

A. Governing equations for the silencer chamber

The acoustic wave equation in region R_2 is given by

$$\frac{1}{c_0^2}\frac{D^2 p_2'}{Dt^2} - \nabla^2 p_2' = 0, \tag{1}$$

where c_0 is the isentropic speed of sound, p' is the acoustic pressure, and t is time. For regions 2 and 3 a coupled modal solution for an axial wave number λ is sought; thus the sound pressure in region 2 is expanded in the form

$$p_2'(x,y,z;t) = p_2(y,z)e^{i(\omega t - k_0 \lambda x)}, \tag{2}$$

where $k_0 (= \omega/c_0)$ is the wave number in region R_2, $i = \sqrt{-1}$, and ω is the radian frequency. Substituting the assumed form for p_2' into the governing wave equation gives

$$\nabla_{yz}^2 p_2 + k_0^2[1 - \lambda M]^2 p_2 - k_0^2 \lambda^2 p_2 = 0, \tag{3}$$

where M is the mean flow Mach number in region R_2 and ∇_{yz} denotes a two-dimensional form of the Laplacian operator (y, z plane).

Similarly for region R_3, if the sound pressure is expanded in the form

$$p_3'(x,y,z;t) = p_3(y,z)e^{i(\omega t - k_0 \lambda x)}, \tag{4}$$

the wave equation may be written as

$$\nabla_{yz}^2 p_3 - (\Gamma^2 + k_0^2 \lambda^2) p_3 = 0, \tag{5}$$

provided mean flow in this region is assumed to be negligible and Γ is the propagation constant of the porous material.

The appropriate boundary conditions that link together regions R_2 and R_3 are the continuity of normal particle displacement and a pressure condition that takes into account the presence of the perforate. It is convenient to write each boundary condition in terms of the acoustic particle velocity; thus for the continuity of displacement,

$$\mathbf{u}_2 \cdot \mathbf{n}_2 = -(1 - M\lambda)\mathbf{u}_3 \cdot \mathbf{n}_3, \quad \text{on } S_c, \tag{6}$$

and for pressure,

$$p_{c_3} - p_{c_2} = \rho_0 c_0 \zeta \mathbf{u}_3 \cdot \mathbf{n}_3, \quad \text{on } S_c. \tag{7}$$

Here, $\mathbf{u}$ is the acoustic velocity vector, $\mathbf{n}$ the outward unit normal vector, and p_c is the sound pressure on boundary S_c (the perforate) either in region R_2 or region R_3. The (dimensionless) acoustic impedance of the perforate is denoted by ζ and ρ_0 is the mean fluid density in region R_2. The assumption of an infinitesimally thin perforate is implicit in Eq. (7) and is valid because the thickness of the perforate is typically small when compared to the overall dimensions of the silencer. Finally, for the outer wall of the silencer chamber (surface S_3), the normal pressure gradient is zero; thus

$$\nabla_{yz} p_3 \cdot \mathbf{n}_3 = 0, \quad \text{on } S_3. \tag{8}$$

B. Finite element discretization and derivation of eigenequation

The acoustic pressure in the chamber is approximated by a trial solution of the form

$$p_2(y,z) = \sum_{J=1}^{N_2} \psi_{2_J}(y,z) p_{2_J}, \tag{9a}$$

and

$$p_3(y,z) = \sum_{J=1}^{N_3} \psi_{3_J}(y,z) p_{3_J}, \tag{9b}$$

for regions R_2 and R_3, respectively. Here $\psi_J(y,z)$ is a global basis function, p_J is the value of $p(y,z)$ at node J, and N_2 and N_3 are the number of nodes in regions R_2 and R_3, respectively. To arrive at the governing eigenequation the weak Galerkin method is adopted and so for region R_2 the wave equation may be rewritten as

$$\int_{R_2} [\nabla_{yz}^2 p_2 + k_0^2[1 - \lambda M]^2 p_2 - k_0^2 \lambda^2 p_2] \psi_I \, dy \, dz = 0, \quad \text{for nodes } I = 1,...,N_2. \tag{10}$$

Applying the Green's theorem to Eq. (10) yields

$$\int_{R_2} [\nabla_{yz} \psi_I \nabla_{yz} p_2 + k_0^2[1 - \lambda M]^2 \psi_I p_2 - k_0^2 \lambda^2 \psi_I p_2] dy \, dz = \int_{S_{c_2}} \psi_I \nabla p_2 \cdot \mathbf{n}_2 \, ds, \tag{11}$$

where S_{c_2} denotes the surface of the perforated tube that lies in region R_2 and s is an element length on surface S_c. Substituting the assumed trial solution for p_2 [Eq. (9a)] into Eq. (11) gives

$$\left\{ \int_{R_2} [\nabla_{yz} \psi_I \nabla_{yz} \psi_J + k_0^2(\lambda^2 - [1 - \lambda M]^2) \psi_I \psi_J] dy \, dz \right\} \{\mathbf{p}_2\} = \int_{S_{c_2}} \psi_I \nabla_{yz} p_2 \cdot \mathbf{n}_2 \, ds. \tag{12}$$

Similarly, the weak Galerkin method allows the wave equation in region R_3 to be written as

$$\left\{ \int_{R_3} [\nabla_{yz} \psi_I \nabla_{yz} \psi_J + (\Gamma^2 + k_0^2 \lambda^2) \psi_I \psi_J] dy \, dz \right\} \{\mathbf{p}_3\} = \int_{S_{c_3}} \psi_I \nabla_{yz} p_3 \cdot \mathbf{n}_3 \, ds, \tag{13}$$

after utilizing the pressure boundary condition on surface S_3 [see Eq. (8)]. Here S_{c_3} denotes the surface of the perforated tube that lies in region 3.

The final eigenequation for the chamber is obtained by using the boundary conditions on the surface S_c to couple together Eqs. (12) and (13). To facilitate the introduction of the pressure and displacement boundary conditions it is necessary first to write the linearized Euler equation, which for regions R_2 and R_3 gives

$$\nabla_{yz} p_2 = -i\rho_0\omega(1-M\lambda)\mathbf{u}_2 \tag{14a}$$

and

$$\nabla_{yz} p_3 = -i\rho(\omega)\omega\mathbf{u}_3, \tag{14b}$$

where $\rho(\omega)$ is the equivalent complex density of the porous material (see Allard and Champoux[13]). By substituting Eq. (14a) onto the right-hand side of Eq. (12), the displacement boundary condition [Eq. (6)] may be introduced, giving

$$\left\{\int_{R_2}[\nabla_{yz}\psi_I\nabla_{yz}\psi_J + k_0^2(\lambda^2-[1-\lambda M]^2)\psi_I\psi_J]dy\,dz\right\}\{\mathbf{p}_2\} = i\rho_0\omega[1-M\lambda]^2\int_{S_{c_2}}\psi_I\mathbf{u}_3\cdot\mathbf{n}_3\,ds. \tag{15}$$

For region 3, the substitution of Eq. (14b) onto the right-hand side of Eq. (13) yields

$$\left\{\int_{R_3}[\nabla_{yz}\psi_I\nabla_{yz}\psi_J + (\Gamma^2+k_0^2\lambda^2)\psi_I\psi_J]dy\,dz\right\}\{\mathbf{p}_3\} = -i\rho(\omega)\omega\int_{S_{c_3}}\psi_I\mathbf{u}_3\cdot\mathbf{n}_3\,ds. \tag{16}$$

The pressure boundary condition [Eq. (7)] may now be substituted onto the right-hand side of both Eqs. (15) and (16) to yield two equations that may then be combined to give a single eigenequation of the form

$$\left\{\int_{R_2}[\nabla_{yz}\psi_I\nabla_{yz}\psi_J + k_0^2(\lambda^2-[1-\lambda M]^2)\psi_I\psi_J]dy\,dz\right\}\{\mathbf{p}_2\} + \left\{\int_{R_3}[\nabla_{yz}\psi_I\nabla_{yz}\psi_J + (\Gamma^2+k_0^2\lambda^2)\psi_I\psi_J]dy\,dz\right\}\{\mathbf{p}_3\} - i\frac{k_0}{\zeta}[1-M\lambda]^2\left\{\int_{S_{C_2}}\psi_I\psi_J ds\right\}\{\mathbf{p}_{C_3}-\mathbf{p}_{C_2}\} + i\frac{k_0}{\zeta}\frac{\rho(\omega)}{\rho_0}\left\{\int_{S_{C_3}}\psi_I\psi_J ds\right\}\{\mathbf{p}_{C_3}-\mathbf{p}_{C_2}\} = 0. \tag{17}$$

Equation (17) constitutes a second-order eigenvalue problem in λ. It is noticeable that the order of this eigenequation has been reduced by 2 when compared to a similar study by Astley and Cummings,[5] who omitted the perforate. Rewriting Eq. (17) in matrix form, and rearranging into ascending orders of λ, gives

$$[[\mathbf{A}]+\lambda[\mathbf{B}]+\lambda^2[\mathbf{C}]]\{\mathbf{p}\} = \{\mathbf{0}\}, \tag{18}$$

where $\mathbf{p}$ is a vector accommodating the pressure in both regions 2 and 3. The matrices $[\mathbf{A}]$, $[\mathbf{B}]$, and $[\mathbf{C}]$ are given by

$$[\mathbf{A}]\{\mathbf{p}\} = [\mathbf{K}_2]\{\mathbf{p}_2\} - k_0^2[\mathbf{M}_2]\{\mathbf{p}_2\} + [\mathbf{K}_3]\{\mathbf{p}_3\} + \Gamma^2[\mathbf{M}_3]\{\mathbf{p}_3\} + \frac{ik_0}{\zeta}[\mathbf{M}_{C_2}]\{\mathbf{p}_{C_2}\} - \frac{ik_0}{\zeta}\frac{\rho(\omega)}{\rho_0}[\mathbf{M}_{C_3}]\{\mathbf{p}_{C_2}\} - \frac{ik_0}{\zeta}[\mathbf{M}_{C_2}]\{\mathbf{p}_{C_3}\} + \frac{ik_0}{\zeta}\frac{\rho(\omega)}{\rho_0}[\mathbf{M}_{C_3}]\{\mathbf{p}_{C_3}\}, \tag{19}$$

$$[\mathbf{B}]\{\mathbf{p}\} = 2Mk_0^2[\mathbf{M}_2]\{\mathbf{p}_2\} - \frac{2ik_0M}{\zeta}[\mathbf{M}_{C_2}]\{\mathbf{p}_{C_2}\} + \frac{2ik_0M}{\zeta}[\mathbf{M}_{C_2}]\{\mathbf{p}_{C_3}\}, \tag{20}$$

$$[\mathbf{C}]\{\mathbf{p}\} = k_0^2(1-M^2)[\mathbf{M}_2]\{\mathbf{p}_2\} + k_0^2[\mathbf{M}_3]\{\mathbf{p}_3\} + \frac{ik_0M^2}{\zeta}[\mathbf{M}_{C_2}]\{\mathbf{p}_{C_2}\} - \frac{ik_0M^2}{\zeta}[\mathbf{M}_{C_2}]\{\mathbf{p}_{C_3}\}, \tag{21}$$

and

$$[\mathbf{K}_2]_{I,J} = \int_{R_2}\nabla_{yz}\psi_I\nabla_{yz}\psi_J\,dy\,dz, \tag{22a}$$

$$[\mathbf{K}_3]_{I,J} = \int_{R_3}\nabla_{yz}\psi_I\nabla_{yz}\psi_J\,dy\,dz, \tag{22b}$$

$$[\mathbf{M}_2]_{I,J} = \int_{R_2}\psi_I\psi_J\,dy\,dz, \tag{22c}$$

$$[\mathbf{M}_3]_{I,J} = \int_{R_3}\psi_I\psi_J\,dy\,dz, \tag{22d}$$

$$[\mathbf{M}_{C_2}]_{I,J} = \int_{S_{C_2}}\psi_I\psi_J\,ds, \tag{22e}$$

$$[\mathbf{M}_{C_3}]_{I,J} = \int_{S_{C_3}}\psi_I\psi_J\,ds. \tag{22f}$$

Finally, the problem may be solved for λ by rewriting Eq. (18) as

$$\begin{bmatrix}\mathbf{0} & \mathbf{I}\\ -[\mathbf{C}]^{-1}[\mathbf{A}] & -[\mathbf{C}]^{-1}[\mathbf{B}]\end{bmatrix}\begin{Bmatrix}\mathbf{p}\\ \lambda\mathbf{p}\end{Bmatrix} = \lambda\begin{Bmatrix}\mathbf{p}\\ \lambda\mathbf{p}\end{Bmatrix}, \tag{23}$$

where $\mathbf{I}$ is an identity matrix.

C. Numerical matching of sound fields

Acoustic pressure and normal particle velocity are to be matched at collocation points on the silencer inlet and exit planes, thus at plane A (see Fig. 1),

$$\mathbf{p}_1'(0,y,z) = \mathbf{p}_2'(0,y,z), \quad (y,z)\in R_2 \ (\text{or } R_1), \tag{24a}$$

$$\mathbf{u}_{x_1}'(0,y,z) = \mathbf{u}_{x_2}'(0,y,z), \quad (y,z)\in R_2 \ (\text{or } R_1), \tag{24b}$$

$$0 = \mathbf{u}_{x_3}'(0,y,z), \quad (y,z)\in R_3, \tag{24c}$$

and for plane B,

$$\mathbf{p}_2'(L,y,z) = \mathbf{p}_4'(L,y,z), \quad (y,z)\in R_2 \ (\text{or } R_4), \tag{25a}$$

$$\mathbf{u}_{x_2}'(L,y,z) = \mathbf{u}_{x_4}'(L,y,z), \quad (y,z)\in R_2 \ (\text{or } R_4), \tag{25b}$$

$$\mathbf{u}_{x_3}'(L,y,z) = 0, \quad (y,z)\in R_3, \tag{25c}$$

where $\mathbf{u}_x'$ is the axial particle velocity. The acoustic pressure and velocity on either side of a discontinuity are now written in terms of a modal expansion, containing both incident and reflected waves. Prior to solving the problem, each modal

expansion must be truncated appropriately. Here the modal sum is truncated at the number of collocation points chosen for an individual region. Thus, in region R_1, if N_1 collocation points are chosen, the sound pressure may be expressed as

$$\mathbf{p}_1'(x,y,z) = P_{i_1}^1 \boldsymbol{\Phi}_i^1 e^{-ik_0 x/(1+M)} + \sum_{n=1}^{N_1} P_{r_1}^n \boldsymbol{\Phi}_r^n e^{-ik_0 \lambda_{r_1}^n x}, \quad (y,z) \in R_1 \ (\text{or } R_2), \tag{26}$$

where $P_{i_1}^1$ is the (known) modal amplitude in the inlet pipe, which is assumed here to contain a plane incident wave only (hence $\boldsymbol{\Phi}_i^1$ is a unit vector of length N_1). Here, the unknown reflected modal amplitudes are denoted by $P_{r_1}^n$, the (known) eigenvalues and associated eigenvectors are denoted $\lambda_{r_1}^n$ and $\boldsymbol{\Phi}_r^n$, respectively, where $\boldsymbol{\Phi}_r$, is a vector of length N_1. Thus, the number of unknown modal amplitudes $P_{r_1}^n$ is equal to the number of collocation points in region R_1. Of course, on applying point collocation it is necessary to map the collocation points in region R_1 onto those in region R_2, and so $N_1 = N_2$. Similarly for region 4,

$$\mathbf{p}_4'(x,y,z) = \sum_{n=1}^{N_4} P_{i_4}^n \boldsymbol{\Phi}_i^n e^{-ik_0 \lambda_{i_4}^n x'}, \quad (y,z) \in R_4 \ (\text{or } R_2), \tag{27}$$

assuming the outlet pipe is terminated anechoically downstream of plane B. Again, the collocation points in region R_4 should map onto those chosen in region R_2, and so $N_2 = N_4$. For the silencer chamber, the overall number of collocation points in regions R_2 and R_3 are chosen as $N_2 + N_3 = N_C$. Hence, the modal expansion of the pressure field in the chamber is given by

$$\mathbf{p}_c'(x,y,z) = \sum_{n=1}^{N_C} P_{i_c}^n \boldsymbol{\Psi}_{i_c}^n e^{-ik_0 \lambda_{i_c}^n x} + \sum_{n=1}^{N_C} P_{r_c}^n \boldsymbol{\Psi}_{r_c}^n e^{-ik_0 \lambda_{r_c}^n x}, \quad (y,z) \in R_2 + R_3, \tag{28}$$

where $P_{i_c}^n$ and $P_{r_c}^n$ are the unknown modal amplitudes for the chamber. For the silencer chamber, the eigenvalues $\lambda_{i_c}^n$ and $\lambda_{r_c}^n$, and the associated eigenvectors $\boldsymbol{\Psi}_{i_c}^n$ and $\boldsymbol{\Psi}_{r_c}^n$ (each of length N_C) are obtained on the solution of Eq. (23). The modal expansions may now be substituted into Eqs. (24) and (25), and the matching conditions enforced at each individual node making up the transverse mesh, thus

$$\sum_{n=1}^{N_2} P_{r_1}^n \boldsymbol{\Phi}_r^n - \sum_{n=1}^{N_c} P_{i_c}^n \boldsymbol{\Psi}_{i_c}^n - \sum_{n=1}^{N_c} P_{r_c}^n \boldsymbol{\Psi}_{r_c}^n = -P_{i_1}^1 \boldsymbol{\Phi}_i^1, \quad \text{on } R_2, \tag{29a}$$

$$\sum_{n=1}^{N_2} P_{r_1}^n \boldsymbol{\Phi}_r^n \frac{\lambda_{r_1}^n}{[1-\lambda_{r_1}^n M]} - \sum_{n=1}^{N_c} P_{i_c}^n \boldsymbol{\Psi}_{i_c}^n \frac{\lambda_{i_c}^n}{[1-\lambda_{i_c}^n M]} - \sum_{n=1}^{N_c} P_{r_c}^n \boldsymbol{\Psi}_{r_c}^n \frac{\lambda_{r_c}^n}{[1-\lambda_{r_c}^n M]} = -P_{i_1}^1 \boldsymbol{\Phi}_i^1, \quad \text{on } R_2, \tag{29b}$$

$$\sum_{n=1}^{N_c} P_{i_c}^n \boldsymbol{\Psi}_{i_c}^n \lambda_{i_c}^n + \sum_{n=1}^{N_c} P_{r_c}^n \boldsymbol{\Psi}_{r_c}^n \lambda_{r_c}^n = 0, \quad \text{on } R_3, \tag{29c}$$

$$\sum_{n=1}^{N_c} P_{i_c}^n \boldsymbol{\Psi}_{i_c}^n e^{-ik_0 \lambda_{i_c}^n L} + \sum_{n=1}^{N_c} P_{r_c}^n \boldsymbol{\Psi}_{r_c}^n e^{-ik_0 \lambda_{r_c}^n L} - \sum_{n=1}^{N_2} P_{i_4}^n \boldsymbol{\Phi}_i^n = 0, \quad \text{on } R_2, \tag{29d}$$

$$\sum_{n=1}^{N_c} P_{i_c}^n \boldsymbol{\Psi}_{i_c}^n \frac{\lambda_{i_c}^n}{[1-M\lambda_{i_c}^n]} e^{-ik_0 \lambda_{i_c}^n L} + \sum_{n=1}^{N_c} P_{r_c}^n \boldsymbol{\Psi}_{r_c}^n \frac{\lambda_{r_c}^n}{[1-M\lambda_{r_c}^n]} e^{-ik_0 \lambda_{r_c}^n L} - \sum_{n=1}^{N_2} P_{i_4}^n \frac{\lambda_{i_1}^n}{[1-M\lambda_{i_1}^n]} \boldsymbol{\Phi}_i^n = 0, \quad \text{on } R_2, \tag{29e}$$

$$\sum_{n=1}^{N_c} P_{i_c}^n \boldsymbol{\Psi}_{i_c}^n \lambda_{i_c}^n e^{-ik_0 \lambda_{i_c}^n L} + \sum_{n=1}^{N_c} P_{r_c}^n \boldsymbol{\Psi}_{r_c}^n \lambda_{r_c}^n e^{-ik_0 \lambda_{r_c}^n L} = 0, \quad \text{on } R_3. \tag{29f}$$

This yields $4N_2 + 2N_3$ equations (the collocation points) and $2N_2 + 2N_c (= 4N_2 + 2N_3)$ unknown modal amplitudes, after putting $P_{i_1}^1 = 1$. Equations (29a)–(29f) may be solved simultaneously to find the unknown modal amplitudes. Finally, the sound transmission loss of the silencer (TL) is given by[11]

$$\text{TL} = -20 \log |p_{i_4}^1|. \tag{30}$$

III. EXPERIMENTAL TESTS

Experimental measurements were performed on two dissipative exhaust silencers, here called silencer A and B. Each silencer is approximately elliptical in cross section and contains a bulk reacting porous material separated from the central airway by a concentric perforated screen (see Fig. 2). The chamber dimensions are summarized in Table I (for each silencer the radius r of the perforated tube is 37 mm.

A. Silencer transmission loss

The silencer transmission loss was measured using the impulse technique described by Cummings and Chang.[11] This method is appropriate in the absence of an anechoic chamber and is suited also to tests that involve mean fluid flow. The technique involves sending a short rectangular pulse through the silencer and capturing the transmitted sound pressure. The process is repeated after a suitable time interval and the transmitted sound pressure successively averaged. The same procedure is followed after removal of the silencer from the test rig and the transmission loss is com-

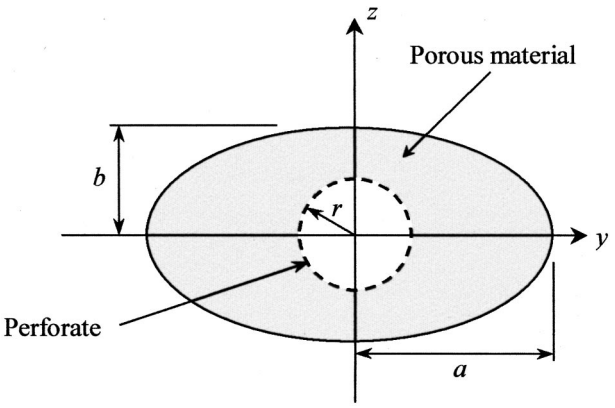


FIG. 2. Dimensions of silencer cross section.

puted by taking the logarithmic ratio of the two captured average sound pressure spectra. A detailed account of the experimental technique is given by the author in a paper on axisymmetric dissipative silencers,[4] although it should be noted here that the impulse technique inevitably incurs frequency limits outside of which experimental measurements are inaccurate. At low frequencies, below approximately 150 Hz, erroneous measurements are common and these are caused largely by reflections from the outlet test pipe arriving back at the silencer before all the reflections within the silencer have died away. An upper frequency limit, approximately 1500 Hz here, is caused by a significant roll-off in the pressure amplitude of the supplied pulse at frequencies above the sampling frequency of 3 kHz.

TABLE I. Data for test silencers.

Silencer	Major axis (a, mm)	Minor axis (b, mm)	Length (L, mm)	Porous material
A	110	60	350	Basalt wool
B	95	50	450	E glass

B. Bulk acoustic properties of the porous materials

Fiber glass and basalt wool are commonly used as acoustic absorbents in automotive silencers. The fiber glass studied here is known commercially in the UK as E glass and has an approximate average fiber diameter of 5–13 μm; the basalt wool studied here has a slightly larger average fiber diameter of 6–18 μm. The analysis in Sec. II demands a knowledge of the bulk complex density $\rho(\omega)$, and the propagation constant Γ for each porous material. It is convenient here to write $\rho(\omega)$ in terms of the (complex) characteristic impedance (z_a), where $\rho(\omega)=z_a\Gamma/i\omega$ (see Allard and Champoux[13]). The propagation constant and characteristic impedance are specified here by combining the empirical power law approach of Delany and Bazley[14] with theoretical low-frequency corrections. The semiempirical approach of Kirby and Cummings[15] alleviates the nonphysical predictions typically obtained when applying Delany and Bazley power laws at low frequencies. For the materials studied here, values for Γ and z_a were given by Kirby and Cummings as

$$\frac{\Gamma}{k_0}=\mathrm{i}\sqrt{\gamma_0 q^2(\omega)}\left\{\frac{[\ln(1-\Omega)+1+2\Omega]\Omega\ln(1-\Omega)+\Omega^2+3\Omega^3/2+\Omega^4/3}{[\ln(1-\Omega)+\Omega+\Omega^2/2]^2}-\left(\frac{\gamma_0-1}{\gamma_0}\right)\mathrm{Pr}-\mathrm{i}\frac{\Omega}{2\pi\xi_f q_0^2 s^2(\omega)}\right\}^{1/2}, \tag{31}$$

$$\frac{z_a}{\rho_0 c_0}=\sqrt{\frac{q^2(\omega)}{\gamma_0\Omega^2}}\left\{\frac{[\ln(1-\Omega)+1+2\Omega]\Omega\ln(1-\Omega)+\Omega^2+3\Omega^3/2+\Omega^4/3}{[\ln(1-\Omega)+\Omega+\Omega^2/2]^2}-\left(\frac{\gamma_0-1}{\gamma_0}\right)\mathrm{Pr}-\mathrm{i}\frac{\Omega}{2\pi\xi_f q_0^2 s^2(\omega)}\right\}^{1/2}, \tag{32}$$

where Ω is the porosity of the porous material, ξ_f is a dimensionless frequency parameter ($\xi_f=\rho_0 f/\sigma_b$, where f is frequency and σ_b is the flow resistivity of the bulk porous material), γ_0 is the ratio of specific heats for air, Pr is the Prandtl number, and q_0 is the so-called steady flow tortuosity. Kirby and Cummings define a ''dynamic'' tortuosity $q^2(\omega)$ and shape factor $s^2(\omega)$ as

$$q^2(\omega)=\frac{[(1+a_3\xi_f^{a_4})(1+a_5\xi_f^{a_6})-a_1a_7\xi_f^{(a_2+a_8)}][\ln(1-\Omega)+\Omega+\Omega^2/2]^2}{[\ln(1-\Omega)+1+2\Omega]\ln(1-\Omega)+\Omega+3\Omega^2/2+\Omega^3/3}, \tag{33}$$

$$s^2(\omega)=\frac{q^2(\omega)}{2\pi\xi_f q_0^2}\frac{1}{[a_1\xi_f^{a_2}(1+a_5\xi_f^{a_6})+a_7\xi_f^{a_8}(1+a_3\xi_f^{a_4})]}, \tag{34}$$

where $a_1\cdots a_8$ are the Delany and Bazley coefficients measured experimentally.[13] The material constants measured for E glass and basalt wool are listed in Table II. Table II also defines a transition value for ξ_f, denoted here ξ_{f_0}, below which $q^2(\omega)$ must be set equal to q_0^2 in Eqs. (31) and (32) (see Kirby and Cummings[15]).

C. Acoustic impedance of perforate screen

The perforate screen used in each of the test silencers was constructed by forming a flat plate with circular perforations into a concentric screen. The acoustic impedance of a perforated plate was shown by Kirby and Cummings[16] to increase when backed by a porous material. They suggested

TABLE II. Porous material constants.

Constant	E glass	Basal wool
a_1	0.2202	0.2178
a_2	−0.5850	−0.6051
a_3	0.2010	0.1281
a_4	−0.5829	−0.6746
a_5	0.0954	0.0599
a_6	−0.6687	−0.7664
a_7	0.1689	0.1376
a_8	−0.5707	−0.6276
σ_b (MKS rayl/m)	30716	13813
Ω	0.952	0.957
q_0^2	5.49	2.91
ξ_{f_0}	0.005	0.0079

the following semiempirical relationship for the nondimensional perforate impedance (ζ),

$$\zeta=\frac{1}{\sigma}\left[\zeta'-\mathrm{i}0.425k_0d+\frac{0.425\,dz_a\,\Gamma}{\rho_0c_0}\right], \tag{35}$$

where d is the diameter of the hole, σ is the area porosity of the perforate, and ζ' is the orifice impedance measured experimentally in the absence of a porous backing, and may be written as $\zeta'=\theta+\mathrm{i}\chi$, where θ is the orifice resistance and χ is the orifice reactance. In the presence of mean flow, Kirby and Cummings[16] proposed the following empirical relationship:

$$\theta=\left(26.16\left[\frac{t}{d}\right]^{-0.169}-20\right)\frac{u_*}{c_0}-0.6537k_0d+\frac{t}{d}\sqrt{8k_0\nu/c_0}, \tag{36}$$

where t is the thickness of the plate, ν is the kinematic viscosity of the mean gas flow, and u_* is the friction velocity of the mean gas flow measured on the inner wall of the perforate. The orifice reactance is given by $\chi=\mathrm{i}k_0(\delta+t)$, and Kirby and Cummings[16] proposed

$$\begin{aligned}&\frac{\delta}{\delta_0}=1, \quad u_*/ft\leqslant 0.18d/t,\\ &\frac{\delta}{\delta_0}=\left(1+0.6\frac{t}{d}\right)\exp\left\{-\frac{u_*/ft-0.18d/t}{1.8+t/d}\right\}\\ &\qquad -0.6t/d, \quad u_*/ft>0.18d/t,\end{aligned} \tag{37}$$

and $\delta_0=0.849d$. When no mean flow is present, θ and χ were given by Bauer[17] as

$$\theta=(1+t/d)\sqrt{8k_0\nu/c_0}, \tag{38a}$$

and

$$\chi=\mathrm{i}k_0(0.25d+t). \tag{38b}$$

IV. RESULTS AND DISCUSSION

The finite element mesh generated for both silencers A and B (see Table I) consisted of six-noded triangular (in region R_2) and eight-noded quadrilateral (in region R_3) isoparametric elements. For both silencers, 24 elements (88 nodes) were used to mesh the chamber, this equates to 35 nodes in region R_2 and 53 nodes in region R_3. Note that in order to implement the pressure change boundary condition across the perforate [see Eq. (7)] it is necessary to place a node on either side of the perforate and then to apply the boundary condition between these two nodes. Thus, the finite element mesh includes two nodes (with identical geometrical coordinates) at each nodal point along the boundary S_c. For the inlet and outlet pipes (regions R_1 and R_4) a mesh identical to the one in region R_2 is used; this facilitates the straightforward application of the point collocation matching technique. The silencer transmission loss is calculated from Eq. (30) after first solving simultaneously Eqs. (29). Transmission loss predictions are compared with experimental measurements in Figs. 3–6, for silencers A and B with mean flow Mach numbers of $M=0$ and $M=0.15$. Theoretical predictions for $M=0$ were obtained by eliminating matrix [**B**] from Eq. (23) prior to solution. For each silencer a concentric perforate screen of thickness $t=1$ mm, hole diameter of $d=3.5$ mm, and an open area porosity of $\sigma=0.263$ was used. When a mean flow Mach number of $M=0.15$ is present, the friction velocity was measured to be $u_*=2.56$ m/s.

It is evident in Figs. 3–6 that good agreement generally exists between measured and predicted silencer transmission loss. For frequencies below 1 kHz, predictions lie within approximately 2 dB of measured values, although the trans-

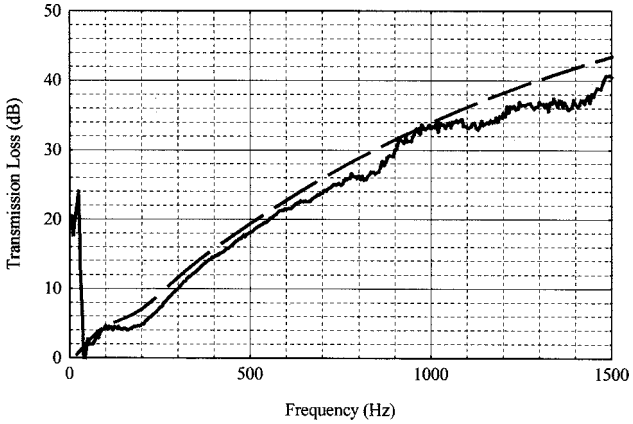


FIG. 3. Transmission loss for silencer A with $M=0$. —, Experimental measurement; – – –, prediction.

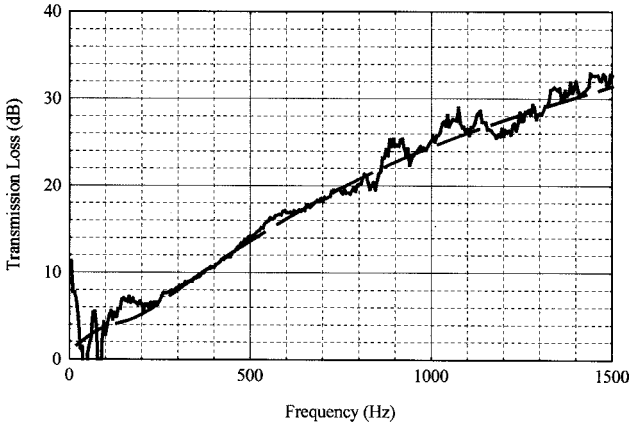


FIG. 4. Transmission loss for silencer A with $M=0.15$. —, Experimental measurement; – – –, prediction.

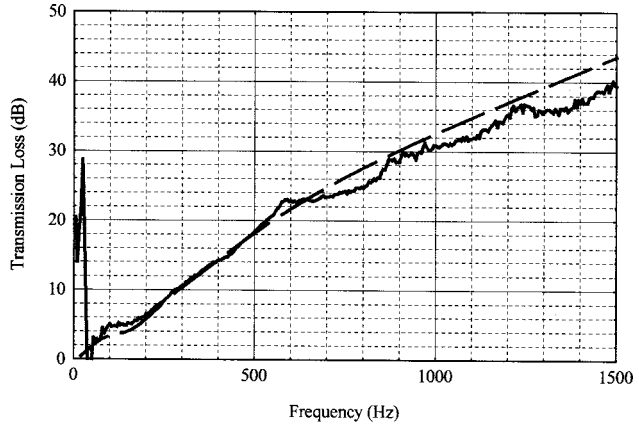


FIG. 5. Transmission loss for silencer B with $M=0$. —, Experimental measurement; – – –, prediction.

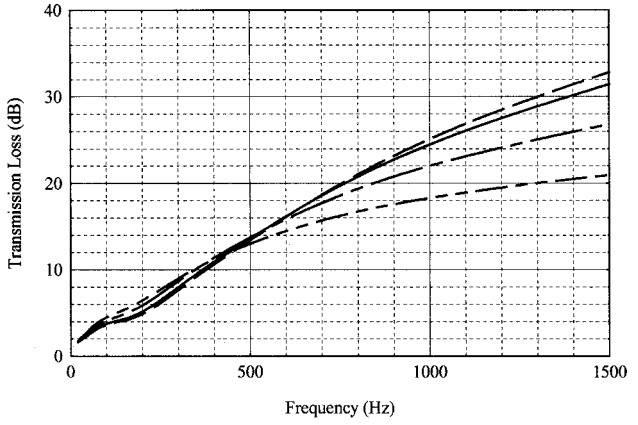


FIG. 7. Transmission loss predictions for silencer A with $M=0.15$. – – –, $\sigma=0.5$; —, $\sigma=0.263$; – – –, $\sigma=0.1$; – – – –, $\sigma=0.05$.

mission loss does tend to be overpredicted. Above 1 kHz, a comparison with experiment is generally less successful and differences of up to 6 dB are evident, largely for those measurements taken without mean flow. Larger discrepancies at higher frequencies are, however, likely to be caused by experimental error. Nevertheless, over the frequency range studied here, the agreement between prediction and experiment is deemed to be acceptable and is at least comparable in accuracy to studies of dissipative silencers by other authors; see, for example, Cummings and Chang,[11] Astley and Cummings,[3] and Aurégan *et al.*[18] Finally, the influence of the perforate is examined in Figs. 7 and 8. Here a number of different values for perforate porosity are examined for silencers A and B with a mean flow Mach number of 0.15. It is evident that, at least for the silencers studied here, a small increase in transmission loss is obtained at low frequencies when the perforate porosity is reduced, however, this is at the expense of a large reduction in transmission loss at higher frequencies.

The computation of silencer transmission loss requires the inversion of three matrices: one of order N_2 (the inlet/outlet eigenvalues), one of order N_c (the chamber eigenvalues), and one of order $2(N_2+N_c)$ (numerical matching). If one assumes a solver speed proportional to N^3 then this CPU expenditure is generally higher than for an equivalent analytic matching procedure. However, in the absence of a reliable analytic mode matching scheme when mean flow is present, CPU expenditure compares favorably with the alternative fully three-dimensional treatment of Peat and Rathi.[1] For example, the method of Peat and Rathi was applied to silencers A and B (see Table I) by Kirby[2] (after omitting the perforate) and, although generally good agreement between the prediction and experiment was observed, the three-dimensional mesh required 1497 degrees of freedom for silencer A and 1757 for silencer B. Thus, for the silencers studied here point collocation represents a considerable saving in CPU run time, amounting to approximately 99.5% for a solver whose speed is proportional to N^3.

The solution of Eq. (23) generates an unordered list of N_c incident and N_c reflected eigenvalues and associated eigenvectors. The imaginary part of the incident and reflected eigenmodes are then sorted into ascending order prior to the application of mode matching. It is noticeable that, on solving Eq. (23) with mean flow present, so-called hydrodynamic modes are not found. This is in contrast to a similar study by Astley and Cummings[5] and is caused by a change in the boundary condition between the airway and the porous material. In the analysis presented here only transverse acoustic particle displacement across the perforate is allowed. More importantly, transverse mean flow effects are suppressed at the perforate that effectively prevents the for-

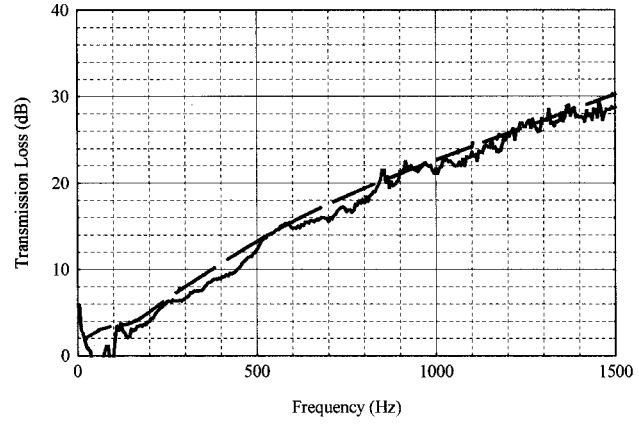


FIG. 6. Transmission loss for silencer B with $M=0.15$. —, Experimental measurement; – – –, prediction.

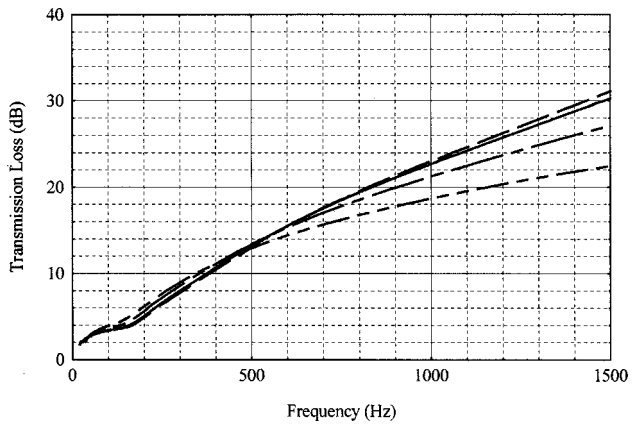


FIG. 8. Transmission loss predictions for silencer B with $M=0.15$. – – –, $\sigma=0.5$; —, $\sigma=0.263$; – – –, $\sigma=0.1$; – – – –, $\sigma=0.05$.

mation of a fully free shear layer and serves to suppress the generation of hydrodynamic modes. Of course, hydrodynamic modes have long been known to exist when mean flow is present, and they are known to play an important role in the performance of reactive silencers, however, when a porous material is present it is likely that strong damping is provided by the material, and this will serve to reduce significantly the influence of hydrodynamic modes on the sound pressure field. Thus, it is assumed in the analysis presented in Sec. II, and the results presented here, that the acoustic performance of the silencer is dominated by the behavior of the least attenuated propagating modes and that the effects of hydrodynamic modes may be neglected.

The transmission loss predictions shown in Figs. 3–8 were obtained after first establishing a converged solution, that is, the number of collocation points were increased until the variation in transmission loss was negligible over the frequency range shown. For convenience, the collocation points were chosen to be identical to the nodal locations chosen for the eigenvalue analysis, and so adaptation of the collocation points effectively takes place prior to carrying out the eigenvalue analysis. Of course, as long as no more than N_c collocation points are chosen, the collocation points need not be coincident with the nodal locations in the original mesh. Moreover, in principle, it should be possible to reduce the size of the collocation problem by reducing the number of collocation points to less than N_c. However, as we do not know, the shape of the final sound pressure field prior to numerical matching, the choice of where best to put the collocation points becomes problematic, especially at higher frequencies. The author has found the most reliable and robust approach by choosing collocation points coincident in location with the nodes chosen for the eigenvalue problem and to adapt the eigenvalue mesh only, i.e., to follow a standard finite element adaptive procedure. Thus, for the current problem, matching is carried out over 35 collocation points in region R_2, and 53 points in region R_3—collocation points equivalent in number to the number of eigensolutions found in the silencer chamber. Of course, it is widely known that for a finite element eigenvalue analysis one can rely on the accuracy only of about 20% of the eigensolutions found. This does not present a problem here since the sound pressure field at an individual collocation point is expressed as a modal sum [see Eqs. (26)–(28)], and it is likely that, at least for the dissipative silencers studied here, the performance of the silencer is dominated by the least attenuated modes, and these are the modes that are found with the most accuracy using the FEM. For the current problem, approximately 18 least attenuated modes may be deemed to be accurate and this number should be more than sufficient to achieve a convergent sum in Eqs. (26)–(28), assuming that hydrodynamic modes may be neglected. Thus, to solve the problem efficiently one needs sufficient least attenuated modes to represent accurately the sound pressure field at an individual collocation point, but also a sufficient number of points to accommodate the variation in the transverse sound pressure field. For the current problem, 88 collocation points are chosen in the chamber and, to maintain a square matrix and hence a tractable problem, 88 eigenmodes are used in Eq. (28). On solving Eqs. (29), errors will therefore be present in 80% of the modal amplitudes found, however, these are amplitudes of highly attenuated modes and so the effect on the overall silencer performance is negligible.

The transmission loss predictions presented in Figs. 3–8 cover a frequency range restricted by experimental measurements. The technique presented here may, however, be used over a much wider frequency range, provided the finite element mesh is adapted in the normal way. This has been demonstrated by Kirby and Lawrie,[12] who successfully applied point collocation to a relatively large HVAC silencer, although mean flow was omitted. Kirby and Lawrie computed the transmission loss for a rectangular duct of cross sectional dimensions 1.5 m×1.5 m and lined on opposite walls with a bulk reacting material. Excellent agreement between point collocation predictions and an exact analytic solution was reported for frequencies up to 2 kHz, providing supporting evidence as to the accuracy of the current method over a wider frequency range than the one presented here (based on a representative k_0a value—"a" being a representative dimension of the duct).

V. CONCLUSIONS

A finite length dissipative silencer of an arbitrary, but uniform, cross section has been modeled by combining a finite element eigenvalue analysis with a point collocation matching scheme. The method is computationally efficient when compared to a three-dimensional finite element approach and avoids the question of modal orthogonality. A good correlation between the prediction and experiment is observed both with and without mean flow, up to a frequency of 1500 Hz for the silencers studied here, although in principle the method is applicable over a much wider frequency range. Furthermore the flexibility and robustness of the finite element method allows the technique to be applied to any cross sectional dissipative silencer geometry, such as rectangular air conditioning ducts, and, in principle, to include any number of duct discontinuities.

[1] K. S. Peat and K. L. Rathi, "A finite element analysis of the convected acoustic wave motion in dissipative silencers," J. Sound Vib. **184**, 529–545 (1995).

[2] R. Kirby, "The acoustic modelling of dissipative elements in automotive exhausts," Ph.D. thesis, University of Hull, UK, 1996.

[3] R. J. Astley, A. Cummings, and N. Sormaz, "A finite element scheme for acoustic propagation in flexible-walled ducts with bulk-reacting liners, and comparison with experiment," J. Sound Vib. **150**, 119–138 (1991).

[4] R. Kirby, "Simplified techniques for predicting the transmission loss of a circular dissipative silencer," J. Sound Vib. **243**, 403–426 (2001).

[5] R. J. Astley and A. Cummings, "A finite element scheme for attenuation in ducts lined with porous material: comparison with experiment," J. Sound Vib. **116**, 239–263 (1987).

[6] K. L. Rathi, "Finite element acoustic analysis of absorption silencers with mean flow," Ph.D. thesis, Loughborough University, UK, 1994.

[7] A. Cummings, "A segmented Rayleigh–Ritz method for predicting sound transmission in a dissipative exhaust silencer of arbitrary cross section," J. Sound Vib. **187**, 23–37 (1995).

[8] R. Glav, "The point-matching method on dissipative silencers of arbitrary cross section," J. Sound Vib. **189**, 123–135 (1996).

[9] M. Åbom, "Derivation of four-pole parameters including higher order mode effects for expansion chamber mufflers with extended inlet and outlet," J. Sound Vib. **137**, 403–418 (1990).

[10] R. Glav, "The transfer matrix for a dissipative silencer of arbitrary cross section," J. Sound Vib. **236**, 575–594 (2000).

[11] A. Cummings and I.-J. Chang, ''Sound attenuation of a finite length dissipative flow duct silencer with internal mean flow in the absorbent,'' J. Sound Vib. **127**, 1–17 (1988).
[12] R. Kirby and J. B. Lawrie, ''Modelling dissipative silencers in HVAC ducts,'' *Proceedings of the Institute of Acoustics Spring Conference*, Salford University, Salford, UK, 2002, Vol. 24, Part 2.
[13] J. F. Allard and Y. Champoux, ''New empirical equations for sound propagation in rigid frame fibrous materials,'' J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[14] M. E. Delany and E. N. Bazley, ''Acoustical properties of fibrous materials,'' Appl. Acoust. **3**, 105–116 (1970).
[15] R. Kirby and A. Cummings, ''Prediction of the bulk acoustic properties of fibrous materials at low frequencies,'' Appl. Acoust. **56**, 101–125 (1999).
[16] R. Kirby and A. Cummings, ''The impedance of perforated plates subjected to grazing gas flow and backed by porous media,'' J. Sound Vib. **217**, 619–636 (1998).
[17] A. B. Bauer, ''Impedance theory and measurements on porous acoustic liners,'' J. Aircr. **14**, 720–728 (1977).
[18] Y. Aurégan, A. Debray, and R. Starobinski, ''Low frequency sound propagation in a coaxial cylindrical duct: application to sudden area expansions and to dissipative silencers,'' J. Sound Vib. **243**, 461–473 (2001).

Improving robustness of active noise control in ducts

Jing Yuan[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kwoloon, Hong Kong, People's Republic of China

(Received 5 July 2002; revised 30 March 2003; accepted 14 April 2003)

A robust active noise controller (ANC) is proposed here for finite ducts. While the $\mathbf{H}_\infty$ control theory provides theoretical ground and numerical algorithms to design robust controllers, it is important for an engineer to design and formulate a robust controller so that the objective is more achievable and the $\mathbf{H}_\infty$ constraints less restrictive without sacrificing robustness. A new robust ANC is designed this way with an extra actuator to improve achievable performance and introduce more degrees of freedom to controller parameters. The new strategy relaxes $\mathbf{H}_\infty$ constraints without sacrificing robustness and enables the ANC to tolerate a wide variety of errors and uncertainties including truncation errors between a finite model and an infinite field. Theoretical analysis, numerical examples, and experimental results are presented to demonstrate the improved performance of the proposed ANC when subject to a certain level of uncertainties in a duct. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1579005]

PACS numbers: 43.50.Ki [KAC]

I. INTRODUCTION

ANC has been proven by many researchers as a viable method[1,2] for noise suppression in low frequency ranges where traditional passive noise control devices become massy, bulky, or less effective. Many ANC schemes synthesize feedforward signals to suppress noise at measured locations. A feedforward ANC usually depends on an accurate secondary path model to adjust its control transfer function.[3–5] Feedback ANC schemes, if properly designed, are able to introduce active damping to sound fields.[6] Hybrid feedforward and feedback ANC is an active research topic.[7–9] Most of the ANC schemes are based on an assumption that acoustical paths are linear time invariant.

In reality, acoustical paths, while linear, are not necessarily time invariant. Online variance of acoustical paths could affect stability of an otherwise stable ANC. Design of robust ANC schemes attracts attentions of many researchers.[10–12] A common practice of robust ANC design is to fit an ANC scheme into the standard four-port framework of robust control theory[13,14] and solve a robust controller subject to a set of $\mathbf{H}_\infty$ constraints. The $\mathbf{H}_\infty$ constraints for robust ANC's are commonly in terms of error bounds of transfer functions, with more weighting outside the control band than the in-band region.[12]

Since parameters of $\mathbf{H}_\infty$ controllers are usually numerical solutions of objective functions subject to certain constraints,[15,16] it is difficult to predict the performance of a robust controller when one designs the structure and formulates the control objective and $\mathbf{H}_\infty$ constraints. Some general design principles would be (1) specifying a control objective as achievable as possible; (2) allowing more DOF in controller parameters; and (3) relaxing, wherever possible, the $\mathbf{H}_\infty$ constraints without sacrificing robustness. This article will show that there is room to improve a robust ANC by adopting the above principles.

The first improvement is related to nonminimum phase (NMP) transfer functions from actuators to usually uncollocated sensors in ANC applications. These transfer functions do not have stable inverses. The use of pseudoinverses prevents ''perfect'' ANC cancellation. It may complicate the ''waterbed effect'' in some robust feedback ANC.[17] While there is no complete solution to the NMP problem for conventional ANC structures, perfect ANC cancellation is analytically possible by an extra actuator.[18] More importantly, the extra actuator introduces more DOF to improve robustness as to be shown here. This is a new feature of the proposed robust ANC and it does improve performance as verified by experiment and numerical results.

It is well known that transfer functions in lightly damped enclosures are rational with numerator and denominator polynomials. There are many sophisticated algorithms identifying rational transfer functions with solid bounds on the parameter errors.[19] For the same level of uncertainties in a lightly damped finite duct, it can be shown, analytically and experimentally, that the $\mathbf{H}_\infty$ constraints are less restrictive if expressed in terms of error bounds of numerators and denominators instead of error bounds of transfer functions. A new formulation is derived for robust ANC in finite ducts. The resultant ANC has an improved experimental performance. This is another improvement over existing robust ANC schemes.

For a hybrid ANC, it is important to design a finite dimensional feedback part to control an infinite dimensional acoustical field without spillover. There are important results on spillover effects and truncation errors.[20–22] This study improves the available results by including actuator/sensor dynamics in the model and tolerating other errors, such as online variance of path parameters. The new ANC achieves this objective using error polynomials to represent a wide variety of uncertainties whose effects may be within the frequency range of interest. This is a more practical method for a control engineer with little background in acoustics to design

[a)]Electronic mail: mmjyuan@polyu.edu.hk

 0001-4966/2003/114(1)/210/8/$19.00

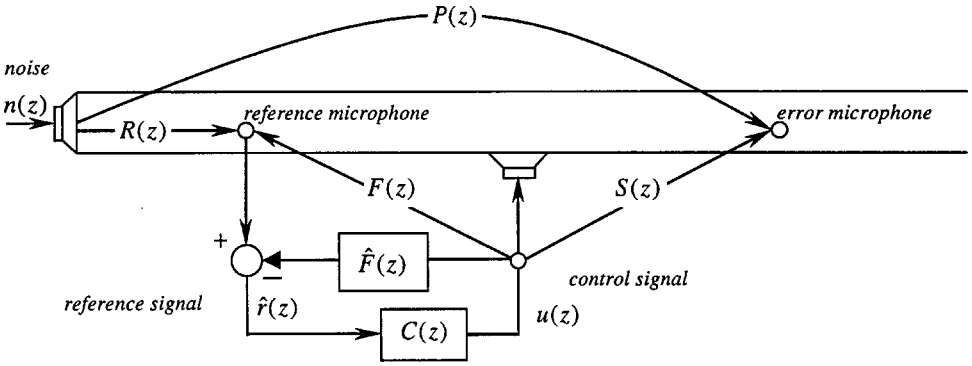


FIG. 1. A typical feedforward ANC system.

ANC. Its advantages are explained here with experimental and numerical verifications.

II. ANALYTICAL GROUNDS OF THE PROPOSED ANC

Figure 1 plots a typical feedforward ANC system in a finite duct, with a primary source, a secondary source, an upstream reference sensor, and a downstream error sensor. The acoustical paths are modeled by $P(z)$, $R(z)$, $F(z)$, and $S(z)$ for the primary, reference, acoustical feedback, and secondary paths, respectively. The reference signal $r(z)$ is mixed with acoustical feedback when the ANC is active. A common way to cancel acoustical feedback is to subtract $\hat{F}(z)u(z)$ from the measured signal, where $\hat{F}(z)$ is an identified version of $F(z)$. As the result, one can express the error signal $e(z)$ as

$$e(z)=\left[P(z)+\frac{S(z)C(z)R(z)}{1-\Delta F(z)C(z)}\right]n(z), \tag{1}$$

where $\Delta F(z)=F(z)-\hat{F}(z)$ is the model error of acoustical feedback path. The ANC system is stable if $\|\Delta F(z)C(z)\|_\infty<1$ by the small game theory. A typical approach for robust ANC design[10–12] is to fit $\hat{P}(z)$, $\hat{R}(z)$, and $\hat{S}(z)$ into the four-port framework of the $\mathbf{H}_\infty$ control theory, and solve for a stable $C(z)$ that minimizes $\|\hat{P}(z)+\hat{S}(z)C(z)\hat{R}(z)\|$ subject to $\|\Delta_F C(z)\|_\infty<1$ if the model error $\Delta F(z)$ could be bounded by Δ_F.

A problem with the above approach is the NMP distortion caused by transfer functions from actuators to usually uncollocated sensors in ANC applications. This means an unstable $S^{-1}(z)$ and makes it impossible to solve $\|P(z)+S(z)C(z)R(z)\|=0$ with a stable $C(z)$ even without any constraints. If subject to $\|\Delta_F C(z)\|_\infty<1$, $\|P(z)+S(z)C(z)R(z)\|$ is definitely non-zero and no smaller than what it could be without constraints. It means a sacrifice of performance for robustness when the best achievable performance is an uncertainty in the first place.

It is desirable to improve robustness and performance simultaneously. A possible way is to modify the ANC structure by adding an extra secondary source. According to the modal theory, path transfer functions in a resonant sound field are rational functions with a common denominator,[18] namely

$$P(z)=\frac{n_p(z)}{d(z)},\quad \mathbf{S}(z)=\frac{1}{d(z)}[n_{s1}(z)\quad n_{s2}(z)],$$

and (2a)

$$R(z)=\frac{n_{r1}(z)}{d(z)}.$$

The degrees of polynomials $d(z)$, $n_p(z)$, $n_{s1}(z)$, $n_{s2}(z)$, and $n_r(z)$ are $2m$ and $2m-1$ if the model is truncated to the first m modes. A properly designed diagonal transfer matrix $\mathbf{C}(z)=\mathrm{Diag}[c_1(z),c_2(z)]$ will translate objective function $\|P(z)+\mathbf{S}(z)\mathbf{C}(z)R(z)\|$ into

$$\|n_p(z)R^{-1}(z)+n_{s1}(z)c_1(z)+n_{s2}(z)c_2(z)\| =0 \text{ or a minimum value.} \tag{2b}$$

This is a Bezout equation[13,14] solvable if $n_{s2}(z)$ does not share roots with $n_{s1}(z)$, which is not a problem for most ANC applications.[18] There exists scalar polynomials $c_1(z)$ and $c_2(z)$, with a finite degree $2m$ and bounded coefficients, such that Eq. (2b) equals zero.[18]

For robust ANC, the solution of Eq. (2b) is subject to $\|\Delta_F\mathbf{C}(z)\|_\infty<1$ that means a tradeoff between performance and robustness. The objective function may not necessarily be zero if subject to the $\mathbf{H}_\infty$ constraint. However, since Eq. (2b) has a zero solution in the unconstrained case, it is reasonable to expect its constrained solution be better than what is achievable by the constrained solution of a traditional ANC, as to be shown in this paper. It will also be made clear that the extra actuator improves ANC robustness by allowing more DOF in the controller parameters.

Another way to improve ANC performance is to reduce Δ_F when an ANC is subject to the same environment conditions. Being a bound of uncertainties, Δ_F should bound two types of possible errors. The initial identification error and the potential online variance of $F(z)$. For a rational path transfer function, the input-output relation can be expressed by an autoregressive and moving average (ARMA) model

$$y_t-d_1y_{t-1}-d_2y_{t-2}-\cdots d_py_{t-p} =n_1u_{t-d}+n_2u_{t-d-1}+\cdots n_{p-1}u_{t-d-p+1},$$

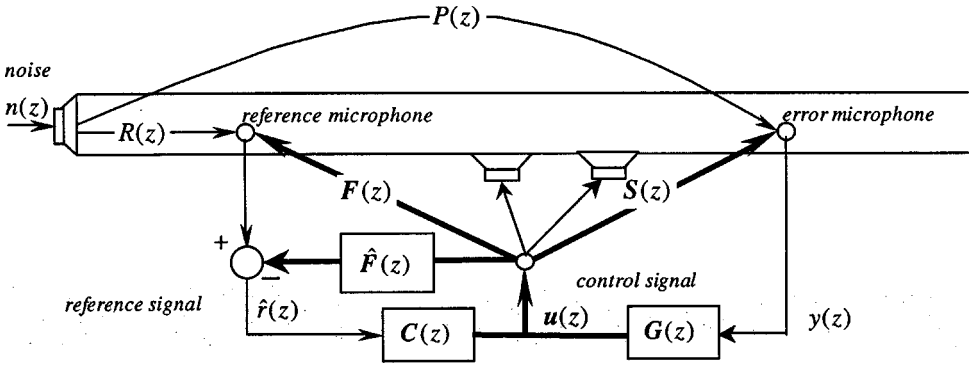


FIG. 2. The proposed ANC system.

where u_t and y_t are the input and output signals of a path model. There are many available algorithms[19] to identify parameters $\{\hat{d}_i\}_{i=1}^{p}$ and $\{\hat{n}_i\}_{i=1}^{p-1}$. Identification errors and online variance of a transfer function may be described by $\Delta d(z)$ and $\Delta n(z)$, respectively.

In theory, acoustical path transfer functions in a duct have infinite degrees. The identified transfer functions, however, can only have numerators and denominators with finite degrees. Uncertainty polynomials $\Delta d(z)$ and $\Delta n(z)$ also describe the effect of model truncation in the design of a robust ANC scheme, especially when the ANC includes a feedback control part. Expressing $\mathbf{H}_\infty$ constraints in terms of error polynomials enables an ANC to tolerate truncation error between a finite model and an infinite field. While the exact forms of $\Delta d(z)$ and $\Delta n(z)$ are not known, it is possible to estimate solid bounds on these polynomials.[19] For a rational transfer function $F(z)=n_f(z)/d(z)$, the error polynomials imply $\Delta F(z)=n_f(z)\Delta d(z)-d(z)\Delta n_f(z)/d^2(z)$ and

$$\Delta_F \propto \left\| \frac{1}{d^2(z)} \right\|_\infty (|\Delta d|+|\Delta n_f|) \tag{3}$$

as a conservative bound for $\Delta F(z)$. Similar bounds can be derived for all paths including secondary paths.

The same level of identification errors $\Delta d(z)$ and $\Delta n(z)$ may have different impact to $\Delta F(z)$ depending on $\|1/d(z)\|_\infty$, as suggested by Eq. (3). For an acoustical field, $d(z)=0$ is the system characteristic equation. This means large $\|\Delta F(z)\|$ or $\|\Delta S(z)\|$ for lightly damped ducts when both $|\Delta d(z)|$ and $|\Delta n(z)|$ are small. If a robust controller can be formulated to tolerate $|\Delta d(z)|$ and $|\Delta n(z)|$ instead of $\|\Delta F(z)\|$ or $\|\Delta S(z)\|$, then the robust constraints are less restrictive to the main objective function like Eq. (2b). This idea leads to a new ANC design strategy in the present study.

If the cross section of a duct is small enough and the duct length is equivalent to a propagation dely of k_l sampling intervals, then it is not difficult to see

$$d(z)=1-\gamma_o\gamma_u(z)\gamma_d(z)z^{-2k_l},$$

where $0<\gamma_o\approx 1$ is the round-trip attenuation, $\gamma_u(z)$ and $\gamma_d(z)$ represent, respectively, the up- and down-stream reflections of the duct. The resonant condition is characterized by a round-trip phase condition $\gamma_o\gamma_u(e^{-j\omega_i})\times\gamma_d(e^{-j\omega_i})e^{-2jk_l\omega_i}=\gamma_i$ with positive real numbers γ_i and $1\leqslant i\leqslant k_l$. The ith resonant peak is proportional to $1/(1-\gamma_i)$. Let $\gamma=\max\{\gamma_i\}$, then $\|1/d(z)\|_\infty\propto 1/(1-\gamma)$. An attempt to reduce Δ_F is linked to the attempt to reduce γ, as suggested by Eq. (3). Since a properly designed feedback controller is able to introduce active damping to a sound field,[6] the proposed ANC includes a pole placement controller. Its objective is to reduce $\|1/d(z)\|_\infty\propto 1/1-\gamma$ by shifting all roots of $d(z)$ into a disc with radius $\gamma_m<\gamma$.

III. ANC DESIGN PROCEDURES

The proposed ANC structure is depicted in Fig. 2, where thick lines represent the paths of vector signals. The actuation signal is a vector synthesized by

$$\mathbf{u}(z)=\mathbf{C}(z)\hat{\mathbf{r}}(z)+\mathbf{G}(z)y(z), \tag{4}$$

where $y(z)$ represents the feedback signal from the error sensor, or optionally plus the signal from the reference sensor. For the sake of simplicity, only the error signal is used here as the feedback. The ANC design can be carried out in two steps. The first step is design of a robust feedback gain $\mathbf{G}(z)$ and the second one is the design of a robust feedforward gain $\mathbf{C}(z)$.

A. Design procedures for the feedback part

For robust pole placement, one may only work on $\mathbf{S}(z)$ without being distracted by other transfer functions (matrices) in the ANC system. Robust pole placement for a path model $\mathbf{S}(z)$ is similar to robust stabilization of a plant $\mathbf{S}(z)$ subject to the same set of $\mathbf{H}_\infty$ constraints. Being a multiple-input–multiple-output (MIMO) transfer matrix, $\mathbf{S}(z)$ has a left- and a right-coprime factorization,[13,14] which are identified from offline data as

$$\mathbf{S}(z)=\frac{1}{\hat{d}(z)}[\hat{n}_{s1}(z)\ \ \hat{n}_{s2}(z)]=\hat{d}^{-1}(z)\hat{\mathbf{N}}_s(z) \tag{5a}$$

and

$$\hat{\mathbf{S}}(z)=[\hat{n}^*_{s1}(z)\quad \hat{n}^*_{s2}(z)]\hat{\mathbf{D}}^{*-1}(z)=\hat{\mathbf{N}}^*_s(z)\hat{\mathbf{D}}^{*-1}(z), \tag{5b}$$

where $\hat{\mathbf{D}}^*(z)$ is a 2×2 polynomial matrix. The coprime factorizations relate to each other by[13,14]

$$\hat{d}(z)\hat{\mathbf{N}}^*_s(z)=\hat{\mathbf{N}}_s(z)\hat{\mathbf{D}}^*(z). \tag{6}$$

There are many sophisticated algorithms to identify these transfer matrices with sufficient accuracy.[19]

As stated previously, robust pole placement is very similar to robust stabilization. There exists a feedback controller[13,14]

$$\mathbf{G}_o(z)=\frac{-1}{a_o(z)}\begin{bmatrix} b_{o1}(z) \\ b_{o1}(z) \end{bmatrix}=\frac{-1}{a_o(z)}\mathbf{B}_o(z)$$

whose elements are scalar polynomials, with finite degrees, solved from a Bezout equation

$$\hat{d}(z)a_o(z)+\hat{n}_{s1}(z)b_{o1}(z)+\hat{n}_{s2}(z)b_{o2}(z)=d_p(z). \quad (7)$$

This controller is able to force the closed-loop characteristic equation to $d_p(z)=0$, but is not necessarily robust with respect to the uncertainties.

An important feature of the proposed ANC is more DOF in its parameters, thanks to the extra actuator. There exists an arbitrary finite-degree polynomial matrix $\boldsymbol{\chi}(z)=[\chi_1(z)\ \chi_2(z)]^T$ to parametrize a family of infinitely many controllers[13,14]

$$\mathbf{G}(z)=\frac{-1}{a_o(z)+\hat{\mathbf{N}}_s^*(z)\boldsymbol{\chi}(z)}[\mathbf{B}_o(z)-\hat{\mathbf{D}}^*(z)\boldsymbol{\chi}(z)]$$

$$=\frac{-1}{a(z)}\mathbf{B}(z). \quad (8)$$

All controllers described by Eq. (8) will place the closed-loop poles as roots of $d_p(z)=0$.[13,14] An ANC designer has ample freedom to choose the most robust controller from the above family.

Due to initial identification errors, environmental variance, and model truncation, the real secondary path transfer matrix is given by $\mathbf{S}(z)=d^{-1}(z)\mathbf{N}_s(z)$ that deviates from Eq. (5a) by

$$d(z)=\hat{d}(z)+\Delta d(z) \quad (9a)$$

and

$$\mathbf{N}_s(z)=\hat{\mathbf{N}}_s(z)+\Delta\mathbf{N}_s(z)$$

$$=[\hat{n}_{s1}(z)+\Delta n_{s1}(z)\quad \hat{n}_{s1}(z)+\Delta n_{s1}(z)]. \quad (9b)$$

When the feedback controller is switched on, the contribution of the secondary sources may be described by

$$\mathbf{S}_c(z)\mathbf{C}(z)\hat{\mathbf{r}}(z)=\frac{1}{1-\mathbf{S}(z)\mathbf{G}(z)}\mathbf{S}(z)\mathbf{C}(z)\hat{\mathbf{r}}(z)$$

in view of Eq. (4). The closed-loop secondary path has a transfer matrix

$$\mathbf{S}_c(z)=\frac{1}{1-\mathbf{S}(z)\mathbf{G}(z)}\mathbf{S}(z)$$

$$=\frac{a(z)}{d_p(z)+\Delta\, da(z)+\Delta N_s B(z)}\mathbf{N}_s(z) \quad (10)$$

upon substitution of Eqs. (5)–(9). If the uncertainties $\Delta d(z)$, $\Delta n_{s1}(z)$, and $\Delta n_{s2}(z)$ are bounded respectively by weighting functions Δ_d, Δ_{s1}, and Δ_{s2}, then one would like to find a controller from the family of Eq. (8) such that (i) roots of $d_p(z)=0$ are in a disc with radius 1; and (ii) the following constraint is satisfied:

$$\left\|\frac{\Delta_d a(z)+\Delta_{s1}b_1(z)+\Delta_{s2}b_2(z)}{d_p(z)}\right\|_\infty<1. \quad (11)$$

If such a controller exists, then the closed-loop poles will remain stable as long as the uncertainties remain bounded by their respective weighting functions.

There are available algorithms[15,16] to solve a robust controller from the family described by Eq. (8). For ANC applications, the open-loop poles of $\mathbf{S}(z)$ are always in a disc with radius $\gamma<1$. The objective of a feedback controller is to place these poles in a smaller disc with radius $\gamma_m<\gamma$. The design conditions are modified to (i) roots of $d_p(z)=0$ are in a disc with radius $\gamma_m<\gamma$, and (ii) the left-hand side of Eq. (11) is smaller than $\gamma_m<\gamma$ with γ_m as small as possible. In view of Eq. (3), weighting functions Δ_d, Δ_{s1}, and Δ_{s2} are generally much smaller than $|\Delta\mathbf{S}(z)|$ or Δ_F, especially for lightly damped ducts. The closed-loop poles will remain inside a disc with radius $\gamma_m<\gamma$ as long as uncertainties $\Delta d(z)$, $\Delta n_{s1}(z)$, and $\Delta n_{s2}(z)$ remain bounded by weighting functions Δ_d, Δ_{s1}, and Δ_{s2}. Robust pole placement design is solvable by available numerical algorithms.[15,16] A experimental and numerical example will be presented to demonstrate the advantages of the proposed feedback control part.

B. What is new in the proposed ANC

As stated in the introduction, many researchers have taken advantage of the potential of hybrid ANC schemes to enhance performance. A possible approach of hybrid and robust ANC design was proposed to fit relevant transfer functions into a standard four-port frame work of $\mathbf{H}_\infty$ control theory.[17] The controller minimizes selective signals in the closed-loop. It is expected that minimization of those signals improves robustness with respect to model uncertainties. Other robust ANC schemes are designed with similar strategies.

The proposed ANC is designed with a different strategy in that (1) the objective of the feedback part is pole placement instead of direct noise suppression; and (2) the $\mathbf{H}_\infty$ constraints are specified with respect to error polynomials. This enables an ANC to tolerate small error polynomials $\Delta d(z)$, $\Delta n_{s1}(z)$, and $\Delta n_{s2}(z)$ that could mean a significant $\Delta S(z)$ in lightly damped ducts in view of Eq. (3). The new strategy and structure enable the feedback controller to place closed-loop poles inside a small disk as long as $\Delta d(z)$, $\Delta n_{s1}(z)$, and $\Delta n_{s2}(z)$ are bounded by weighting functions Δ_d, Δ_{s1}, and Δ_{s2}. This control part will reduce uncertainties in all transfer functions, as to be demonstrated by numerical examples in the Sec. IV.

When a finite dimensional feedback controller, like the proposed one, is applied to an infinite dimensional system like a duct, the problem of model truncation and spillover becomes important. There are methods or stability conditions to deal with spillover or ensure stability with respect to truncation errors.[20–22] The actuator and sensor dynamics, which play important roles in ANC control, are absent in those models. Inevitable errors in parameters of model matrices or

online variance of plant parameters are not considered in Refs. 20–22 that focus on flexible structure with relatively invariant plants. For ANC applications, however, errors caused by online variance of path parameters may be equally, if not more, threatening to ANC stability since their effects may be within the frequency band of interest.

On the other hand, bounds on error polynomials are conservative enough to include the effects of a wide variety of errors in transfer functions identified from input–output data.[19] The actuator/sensor dynamics are integrated in these transfer functions. The proposed feedback controller is subject to a set of $\mathbf{H}_\infty$ constraints given by Eq. (11) that is in terms of bounds on error polynomials. It enables the new ANC to tolerate a wide variety of errors including model truncation errors and errors within the frequency band of interest due to online variation of path parameters. In Sec. IV, the proposed ANC will be shown to work well in a finite duct. It also tolerates a set of conservative parameter variations in simulations. These data show that the new design strategy makes the $\mathbf{H}_\infty$ constraints less restrictive without sacrificing robustness. It is an important feature for robust ANC schemes.

C. Design procedures for the feedforward part

Before designing the feedforward part, the primary source should be shut-off and the feedback part of the ANC should be activated with a control signal

$$\mathbf{u}(z)=\mathbf{G}(z)y(z)+\mathbf{v}(z),$$

where $\mathbf{v}(z)$ is substituted by a pseudorandom noise signal. By measuring sound signals from all the sensors, one should be able to identify $\hat{\mathbf{S}}_c(z)$ and $\hat{\mathbf{F}}_c(z)$, respectively. The subscript "c" indicates that the transfer matrices are "closed-loop" ones when the feedback part of the ANC is switched on. Alternatively, one may shut-off $\mathbf{v}(z)$ and turn on the primary source to identify $\hat{P}_c(z)$ and $\hat{\mathbf{R}}_c(z)$ with $\mathbf{u}(z)=\mathbf{G}(z)y(z)$ active. These transfer functions (matrices) are the design basis for the feedforward part of the proposed ANC scheme.

Starting from this step, $\mathbf{u}(z)$ becomes invisible to the design of $\mathbf{v}(z)=\mathbf{C}(z)r(z)$ though $\mathbf{u}(z)$ remains active and its effects are included in transfer functions/matrices identified by the above process. One may focus on $n(z)$ as the exogenous input and $\mathbf{v}(z)$ as the control input. The proposed ANC structure fits well into the framework of a four-port system for the $\mathbf{H}_\infty$ control theory, with a transfer matrix

$$\hat{\mathbf{T}}_c(z)=\begin{bmatrix}\hat{P}_c(z) & \hat{\mathbf{S}}_c(z)\\ \hat{R}_c(z) & \hat{\mathbf{F}}_c(z)\end{bmatrix}.$$

The feedforward control part is synthesized by $\mathbf{v}(z)=\mathbf{C}(z)r(z)$. Its objective is given in Eq. (2b). In this case, one should substitute scalar polynomials $d_c(z)$, $n_{cp}(z)$, $n_{cs1}(z)$, $n_{cs2}(z)$, and $n_{cr}(z)$ as elements of transfer matrices

$$P_c(z)=\frac{n_{cp}(z)}{d_c(z)},\quad \mathbf{S}_c(z)=\frac{1}{d_c(z)}[n_{cs1}(z)\ \ n_{cs2}(z)],$$

and

$$R_c(z)=\frac{n_{cr}(z)}{d_c(z)}$$

to replace those given by Eq. (2a). The corresponding $\mathbf{H}_\infty$ constraint is $\|\Delta_F C(z)\|_\infty<1$. This problem is solvable by the available $\mathbf{H}_\infty$ controller design algorithms.[15,16]

IV. EXPERIMENTAL AND NUMERICAL VERIFICATION

The proposed ANC is tested in an experimental duct with a length of 2 m and a cross section of 11×15 cm^2. The primary source locates at one end of the duct while the two secondary sources are, respectively, 80 cm and 120 cm away from the primary source. The reference sensor is 20 cm away from the primary source and the error sensor 40 cm away from the outlet. The ANC system has a sampling frequency of 2.4 kHz. The dynamics of actuators, sensors, amplifiers, and anti-alias filters are integrated in the transfer functions without any separate treatment.

An important feature of the proposed ANC is robust reduction of resonant peaks in the path transfer functions via feedback control. The experiment verifies such effect in a wide frequency range as shown in Fig. 3. The feedback control reduces the magnitudes of the resonant peaks of all path transfer functions inside the duct. This reduces Δ_F and $\|\Delta S(z)\|$ in view of Eq. (3). Another feature of the proposed ANC is the use of Eq. (2b) to deal with the NMP secondary path transfer functions for broadband cancellation. The cancellation effect of the proposed ANC is illustrated in Fig. 4 where the noise is suppressed in a wide frequency range, as shown by the gray-dashed and black-solid curves, respectively.

The experimental result demonstrates the ANC performance in the presence of model truncation and identification errors. The ANC is also able to tolerate online variance of path parameters. A numerical study is conducted to demonstrate this feature. A duct with a length of $L=3.4$ m is simulated with the first 10 lightly damped modes in the model. The damping ratios are chosen in such a way that the open-loop poles have magnitudes 0.994 when the transfer functions are converted into discrete-time models. In the simulation, the primary source locates at $x_p=0$; the secondary sources locate at $x_{s1}=0.5$ L and $x_{s2}=0.6$ L, respectively. The reference sensor and the error sensor are placed at $x_f=0.2$ L and $x_e=0.7$ L, respectively.

One way to model the online variance is to let the coefficients of $\Delta d(z)$ and $\Delta n(z)$ be zero-mean random numbers with $\sigma=0.05$ for a transfer function with denominator $d(z)$ and numerator $n(z)$. This test may not be realistic, yet it helps to assess the robustness of the ANC whose $\mathbf{H}_\infty$ constraints are set conservatively to tolerate uncertainties including the variance used here. In order to assess the impact of $\Delta d(z)$ and $\Delta n(z)$ generated this way, a Matlab script is written to compute and plot the frequency responses of a set of $\Delta F_1(z)$'s given by

$$\Delta F_1(z)=\frac{n_{f1}(z)}{d(z)}-\frac{n_{f1}(z)+\Delta n_{f1}(z)}{d(z)+\Delta d(z)}.$$

Each member of this set represents a possible distortion caused by a particular version of the variance. The results are

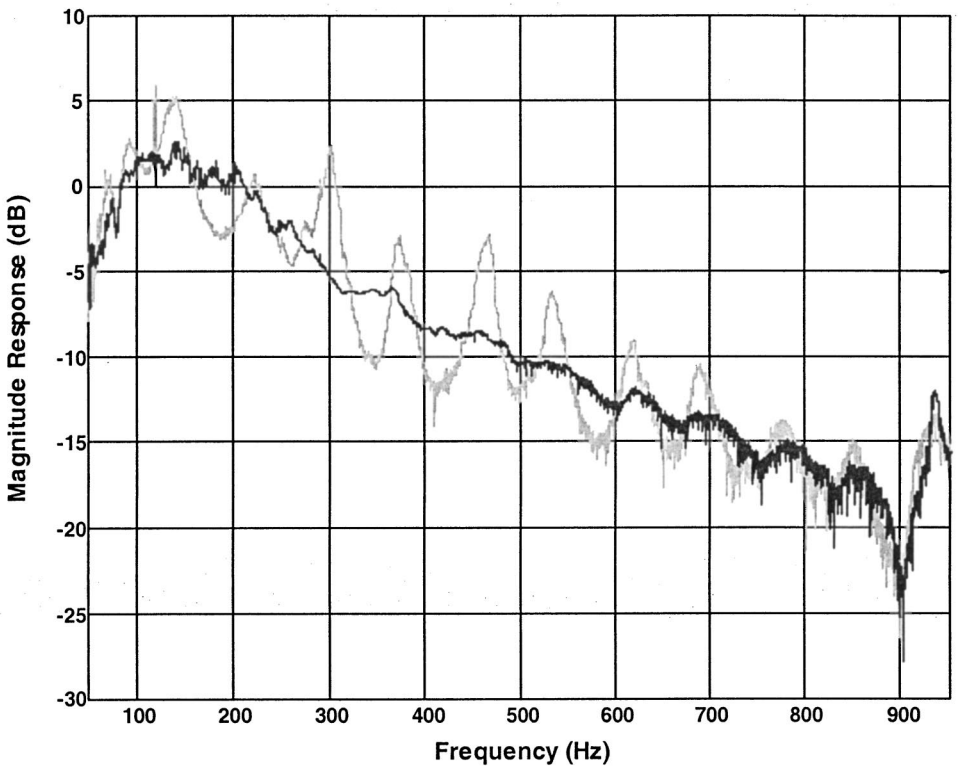


FIG. 3. Magnitude responses of a path from a secondary source to the error sensor with (black-solid) and without (gray-dashed) feedback damping.

shown in Fig. 5 where the lightly damped resonant causes a very small $\|d(z)\|_\infty$ that magnifies the impact of $\Delta d(z)$ and $\Delta n(z)$ to $\|\Delta F_1(z)\|_\infty$.

The effect of feedback damping can be evaluated by comparing any closed-loop transfer function with its open-loop counterpart. For demonstration, $\Delta F_{c1}(z)$ is compared with $\Delta F_1(z)$ when subject to the same level of variances. Frequency responses of a set of $\Delta F_{c1}(z)$'s are plotted in Fig. 6. Each $\Delta F_{c1}(z)$ is computed subject to the same level of variance as $\Delta F_1(z)$ is. Comparing Figs. 5 and 6, one can see that $\|\Delta F_{c1}(z)\|$ is reduced significantly by the feedback control part. The weight function Δ_F can be reduced by 20 dB approximately in the entire frequency of interests.

Since the proposed ANC is subject to a set of less restrictive $\mathbf{H}_\infty$ constraints and does not have the NMP problem, it is expected to have a better cancellation result when sub-

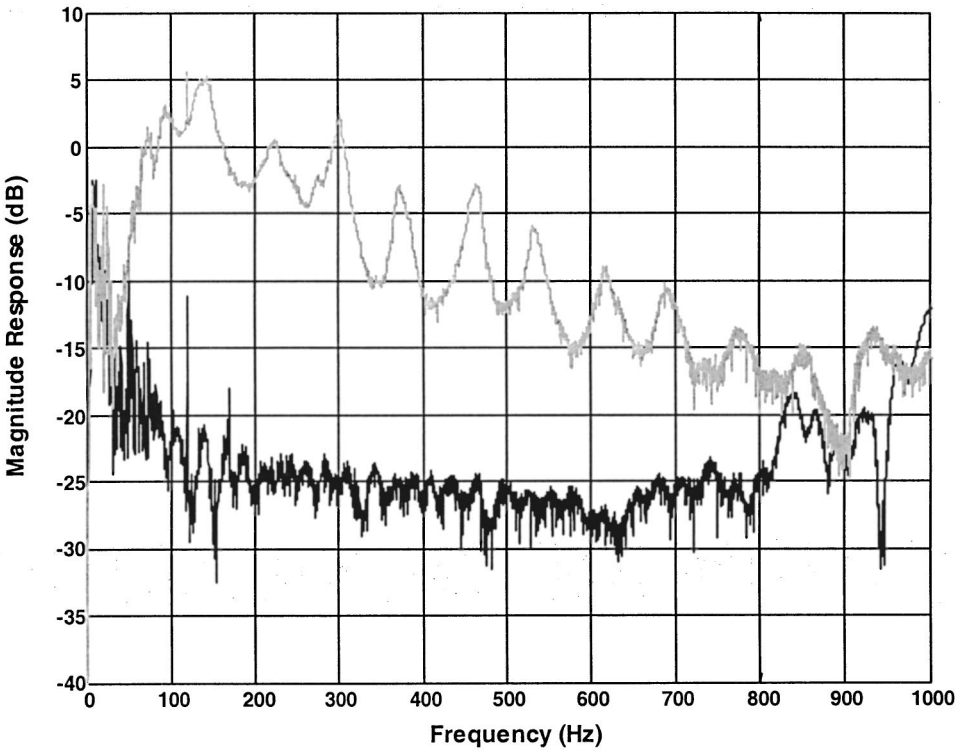


FIG. 4. Magnitude responses of the primary path to the error sensor with (black-solid) and without (gray-dashed) active control.

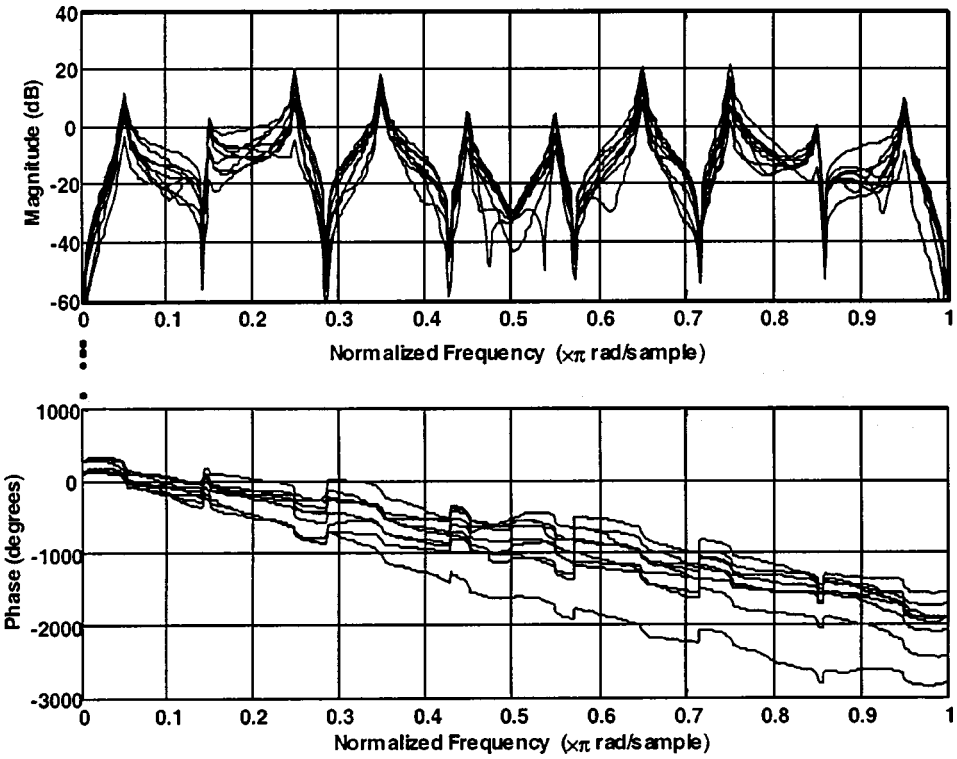


FIG. 5. Frequency responses of a set of $\Delta F_1(z)$'s (without feedback control), when coefficients of $\Delta d(z)$ and $\Delta n_{f1}(z)$ are zero-mean random numbers with $\sigma = 0.05$.

ject to the same level of variances. This is verified by the numerical results shown in Fig. 7. The gray-dashed curve represents the spectrum of a noise without ANC control. The black-solid curves represent the spectra of a set of noise signals with ANC control and subject to a small level of variances in the duct. The ANC is shown able to suppress noise in a wide frequency range subject to the online variance of path parameters.

V. CONCLUSION

A new robust ANC scheme is proposed for finite ducts. It uses an extra actuator to achieve ''perfect cancellation'' of a broadband noise and introduce extra DOF in robust ANC design. Another advantage of the proposed ANC is a new design strategy for its feedback control part. The new structure and the new strategy enable the ANC to tolerate a cer-

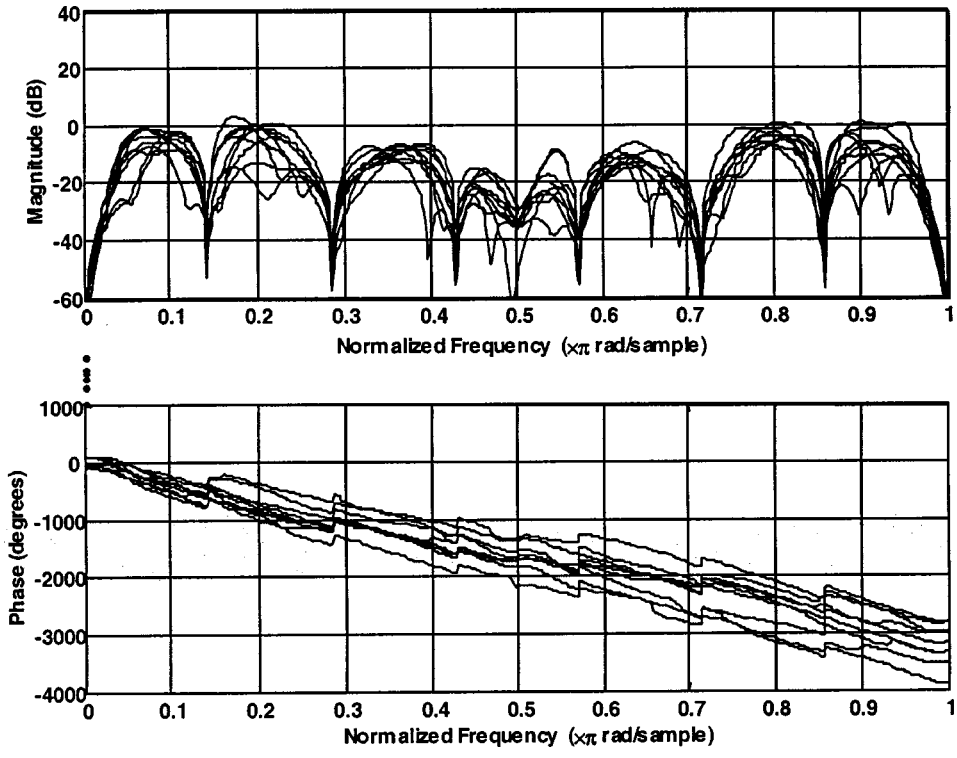


FIG. 6. Frequency responses of a set of $\Delta F_{c1}(z)$'s (with feedback control), when coefficients of $\Delta d(z)$ and $\Delta n_{f1}(z)$ are zero-mean random numbers with $\sigma = 0.05$.

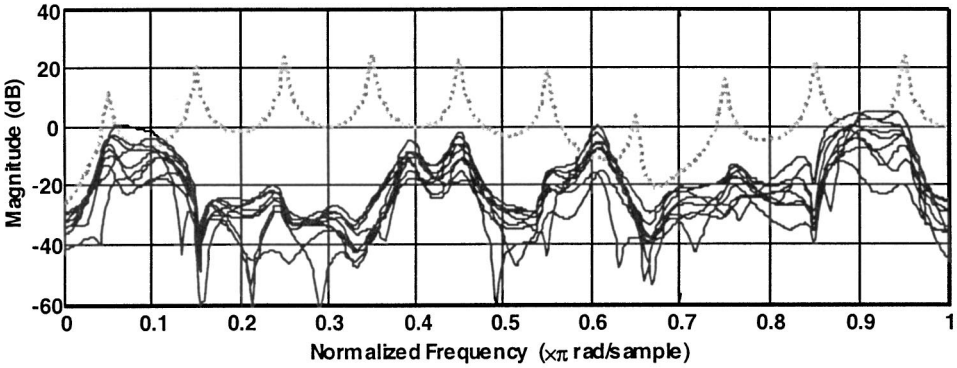


FIG. 7. Cancellation performance of the proposed ANC subject to a small level of variances: a gray dashed curve represents uncontrolled noise spectrum and the thin-black curves represent controlled noise spectra.

tain level of uncertainties and truncation errors to keep closed-loop poles in a disc with a small radius. This reduces the ANC sensitivity to uncertainties in path transfer functions significantly. The proposed ANC system has an improved performance and robustness, verified numerically and experimentally.

ACKNOWLEDGMENT

The work described in this paper was substantially supported by a grant from the Research Grants Council of the Hong Kong Special Administration Region (Project No. PolyU 5175/01E).

[1] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).
[2] C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (E and FN Spon, London, 1997).
[3] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems—Algorithms and DSP Implementations* (Wiley, New York, 1996).
[4] L. J. Eriksson, ''Development of the filtered-U algorithm for active noise control,'' J. Acoust. Soc. Am. **89**, 257–265 (1991).
[5] D. H. Crawford and R. W. Stewart, ''Adaptive IIR filtered-v algorithms for active noise control,'' J. Acoust. Soc. Am. **101**, 2097–2103 (1997).
[6] J. Yuan, ''Global damping of noise or vibration fields with locally synthesized controllers,'' J. Acoust. Soc. Am. **111**, 1726–1733 (2002).
[7] S. J. Elliott and T. J. Sutton, ''Performance of feedforward and feedback systems for active control,'' IEEE Trans. Speech Audio Process. **4**, 214–223 (1996).
[8] W. Tseng, B. Rafaely, and S. J. Elliott, ''Combined feedback-feedforward active control of sound in a room,'' J. Acoust. Soc. Am. **104**, 3417–3425 (1998).
[9] E. Esmailzadeh, A. Alasty, and A. R. Ohadi, ''Hybrid active noise control of a one-dimensional acoustic duct,'' ASME J. Vibr. Acoust. **124**, 10–18 (2002).
[10] B. Rafaely and S. J. Elliott, ''H_2/H_∞ active control of sound in a headset: Design and implementation,'' IEEE Trans. Control Syst. Technol. **7**, 79–84 (1999).
[11] J. Y. Lin and Z. L. Luo, ''Internal model-based LQC/H_∞ design of robust active noise controllers for acoustical duct systems,'' IEEE Trans. Control Syst. Technol. **8**, 864–872 (2000).
[12] M. R. Bai and P. Zeung, ''Synthesis of a robust broadband duct ANC system using convex programming approach,'' J. Acoust. Soc. Am. **111**, 1751–1757 (2002).
[13] B. A. Frances, ''A course in $\mathbf{H}_\infty$ control theory,'' in *Lecture Notes in Control and Information Sciences*, edited by M. Thoma and A. Wyner (Springer-Verlag, New York, 1987).
[14] M. Vidyasagar, *Control System Synthesis: A Factorization Approach* (MIT Press, Cambridge, MA, 1985).
[15] B. M. Chen, *Robust and H_∞ Control* (Springer, London, 2000).
[16] P. Gahinet, A. Nemirovski, A. J. Laub, and M. Chilali, *LMI Control Toolbox* (The MathWorks, Inc., Massachusetts, 1995).
[17] M. R. Bai and H. H. Lin, ''Comparison of active noise control structures in the presence of acoustical feedback by using the $\mathbf{H}_\infty$ synthesis technique,'' J. Sound Vib. **206**, 453–471 (1997).
[18] J. Yuan, ''Relaxed condition for 'perfect' cancellation of broadband noise in 3D enclosures,'' J. Acoust. Soc. Am. **107**, 3235–3244 (2000).
[19] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
[20] M. J. Balas, ''Feedback control of flexible systems,'' IEEE Trans. Autom. Control **23**, 673–679 (1978).
[21] M. J. Balas, ''Trends in large space structure control theory: fondest hopes, wildest dreams,'' IEEE Trans. Autom. Control **27**, 522–535 (1982).
[22] M. J. Balas, ''Stable feedback control of linear distributed parameter systems: time and frequency domain conditions,'' J. Math. Anal. Appl. **225**, 144–167 (1998).

Noise duration for a single overflight

Rufin Makarewicz[a] and Hanna Wojciechowska
Institute of Acoustics, A. Mickiewicz University, 61-614 Poznan, Umultowska 85, Poland

(Received 20 September 2001; revised 14 December 2002; accepted 17 December 2002)

Overflights in national parks and preserves interfere with communication and sounds of nature. The percentage of time that an aircraft is audible, P, can be used as a noise metric. To calculate P the overflight time for a single aircraft, τ, has to be known. The method of τ calculation is based on the assumption that an aircraft is a point source and the noise propagation is governed by geometrical spreading, air absorption, and refraction. The atmosphere is characterized by the effective sound speed gradient. Analytical formulas for τ are derived for down- and crosswind flights. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1553459]

PACS numbers: 43.50.Lj [MRS]

I. INTRODUCTION

Overflights of national parks and preserves by aircraft are a fast growing problem across the world. The noise at low levels from a distant aircraft evokes annoyance not because it interferes with communication, but because it interferes with the natural quiet and sound of nature. How can aircraft activity be kept within acceptable limits? U.S. Public Law 100-91 required an assessment of the noise impact of aircraft overflights. Since the equal-energy hypothesis provides only weak and arguable basis for predicting the response of visitors to recreational areas,[1] a new noise index was proposed, namely, the percentage of time that an aircraft is audible, P.[2,3] This quantity is simply the ratio of the time the visitor records aircraft as audible, $\tau_1+\tau_2+\cdots\tau_n$, to the total time he or she is at the site, T. It is assumed no overlapping of successive aircraft passbys (Fig. 1). Therefore,

$$P=\frac{1}{T}\sum_{i=1}^{n}\tau_i. \tag{1}$$

A similar measure of annoyance of distorted speech was used by Berglund *et al.*[4] In this study the overflight time, τ, for a single aircraft during level flight at constant speed is calculated. It depends on the A-weighted sound level threshold, $\tilde{L}_A$, the parameters describing the flight route and weather conditions. The obtained results could help eventually to design flight routes and minimize the noise annoyance. National parks and preserves are mostly visited under a clear sky during sunny weather. Such weather yields refraction with rays concave upwards. In this study the model for calculating τ is based on geometrical divergence of noise, air absorption, and refraction. The noise impact area is determined by limiting rays which graze the ground tangentially and make a shadow, where no ray enters (Fig. 2).[5] However, some noise energy penetrates into the shadow region due to diffraction as well as scattering by temperature and wind velocity fluctuations.[6–9] These processes limit the reduction of noise levels to between 10 and 30 dB. Furthermore, the shadow boundary is not sharp. The grazing points provide the boundary of the noise impact area. However, noise evokes annoyance when its A-weighted sound level exceeds some threshold, $L_A>\tilde{L}_A$. The common sector of the threshold and the impact area (in the $x-y$ plane) form the annoyance area, where the aircraft noise interferes with the natural quiet and sound of nature. To determine the overflight time, τ, one has to know the impact area (Sec. III) and the threshold area (Sec. IV). To find the former, a prior knowledge of the effective sound speed gradient is necessary (Sec. II).

II. THE EFFECTIVE SOUND SPEED GRADIENT

The boundary layer is a part of the Earth's atmosphere in which the effects of the surface friction and surface heating play a role. During a sunny day, after midday, its maximal thickness equals 1200–1500 m. In this study a flight at an altitude, $H<1500$ m, within the boundary layer is considered. During a sunny day (unstable weather conditions), the boundary layer consists of three regions: the surface layer, the free convection layer, and the mixed layer. The first of them, with nonlinear gradients of air temperature and wind speed, extends a few tens of meters above the ground. Since the flight's altitude equals at least a hundreds meters, the surface layer is relatively thin and can be neglected. The atmosphere above can be described by constant gradients of temperature and wind speed.[10,11] Within the free-convection layer, turbulent flux occurs. Despite that, the vertical component of air motion can be neglected. The atmosphere can be treated as a stratified moving medium, where the wind velocity and temperature depend only on the vertical coordinate z. Making use of least-squares analysis, the mean profiles of the temperature and wind speed above the surface layer can be approximated by linear functions:[10,11]

$$T(z)=T_0-\frac{dT}{dz}z+273.15\ \text{K}, \tag{2}$$

and

$$u=u_0+\frac{du}{dz}z. \tag{3}$$

In Eqs. (2) and (3) the derivatives denote the average temperature and wind speed gradients, respectively. For example: $u_0=12$ m/s, $du/dz=6\times10^{-3}$ 1/s, $T_0=13$ K, and

[a]Electronic mail: makaaku@amu.edu.pl

 0001-4966/2003/114(1)/218/7/$19.00

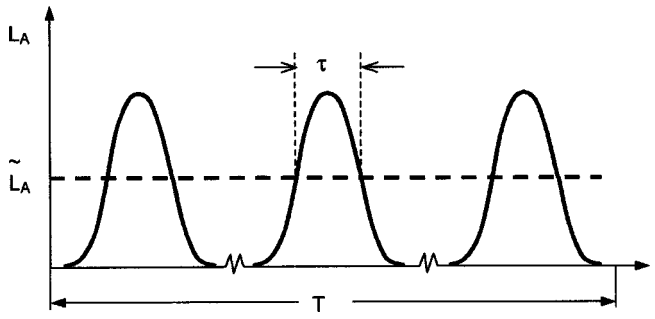


FIG. 1. The A-weighted sound level, L_A, of noise produced by aircraft overflights ($\tilde{L}_A$-threshold level).

$dT/dz = 6\times10^{-3}$ K/m. For a flight altitude of a few hundreds meters, under any weather conditions, the temperature variation is relatively small,

$$T_0 + 273.15 \gg \frac{dT}{dz} z. \tag{4}$$

Since the speed of sound is directly proportional to $\sqrt{T}$,[12] the above-mentioned inequality and Eq. (2) yield the adiabatic sound speed dependence upon altitude,

$$c(z) \approx c(0)[1 - \alpha z], \tag{5}$$

where $c(0) = 331.36 + 0.6067 T_0$ and

$$\alpha = \frac{1}{2(T_0 + 273.15)} \frac{dT}{dz} \tag{6}$$

is the relative temperature gradient. Introducing the relative wind speed gradient,

$$\beta = \frac{1}{c(0)} \frac{du}{dz}, \tag{7}$$

the horizontal angle, Φ, and the angle of the average wind direction, Ψ (Fig. 3), the effective sound speed is,[13–20]

$$c_e(z) \approx c(0)[1 + M\cos(\Phi - \Psi) - \gamma z]. \tag{8}$$

Here $M = u_0/c(0)$ and

$$\gamma = \alpha - \beta\cos(\Phi - \Psi) \tag{9}$$

is the effective sound speed gradient which was introduced first by Weiner and Keast.[21] Note that for $u_0 = 12$ m/s (see above), with the speed of sound at the ground, $c(0)$, exceeding 300 m/s, one gets $M \ll 1$. Similarly, taking $\alpha = 1.1 \times 10^{-5}$ 1/m and $\beta = 1.8\times10^{-5}$ 1/m, for the downwind propagation with $\Phi = \Psi$ one has the maximal value of $\gamma = 2.9\times10^{-5}$ 1/m. For typical weather conditions inequality

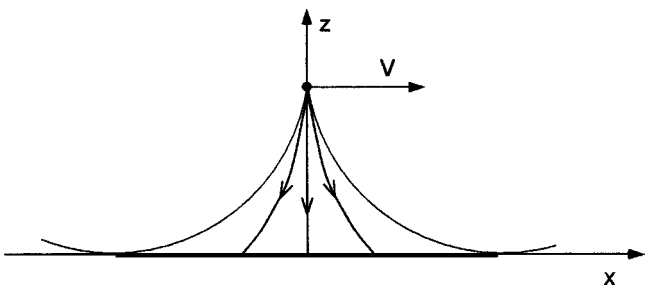


FIG. 2. Cross-section view of noise impact area in the $x-z$ vertical plane.

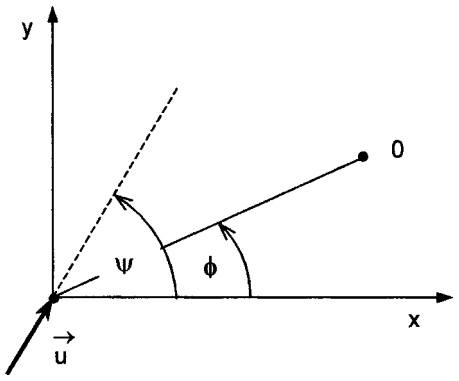


FIG. 3. The horizontal angle, Φ, with the observer O and the angle of wind direction, Ψ.

$\gamma z \ll 1$ holds for the flight altitude $H < 1500$ m which is of interest in this study.

III. IMPACT AREA

Introducing the effective sound speed, the moving atmosphere is replaced by a virtual motionless medium with an adiabatic sound speed equal to c_e [Eq. (8)]. Therefore Snell's law can be applied (Fig. 4):

$$\frac{\sin\Theta(z)}{c_e(z)} = \frac{\sin\Theta(0)}{c_e(0)}. \tag{10}$$

The boundary of the noise impact area is determined by the points on the ground at which the limiting rays become horizontal, $\Theta(0) = \pi/2$. Hence, the limiting ray equation is

$$\frac{dR}{dz} = -\frac{c_e(z)/c_e(0)}{\sqrt{1 - c_e^2(z)/c_e^2(0)}}. \tag{11}$$

In view of Eq. (8), with $M \ll 1$ and $\gamma H \ll 1$ (see above), integration from $z = H$ (aircraft) to $z = 0$ (ground) yields the horizontal distance between the source and the boundary of the impact area (shadow),[22]

$$R(\Phi - \Psi) = R_0\left[1 - \frac{\beta}{\alpha}\cos(\Phi - \Psi)\right]^{-1/2}, \quad 0 < \Phi \leq 2\pi, \tag{12}$$

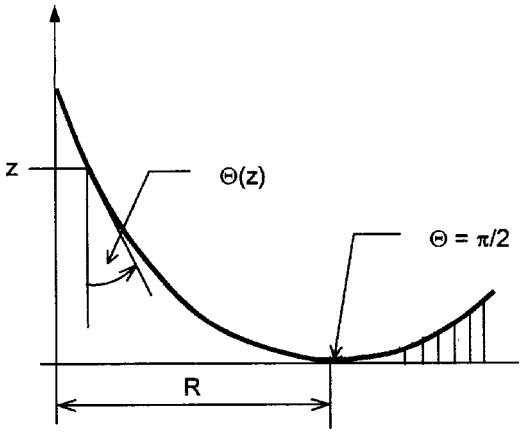


FIG. 4. The limiting ray and the horizontal distance, R [Eq. (12)], between the aircraft and the boundary of the impact area (shadow zone).

where

$$R_0 = \sqrt{2H/\alpha} \tag{13}$$

is the radius of the circular impact area for a windless atmosphere. With the relative temperature gradient, $\alpha = 1.1 \times 10^{-5}$ 1/m, and the altitude, $H = 200$ m, one gets, $R_0 \approx 6$ km.

IV. THRESHOLD AREA

Assuming an aircraft can be modeled by a nondirectional point source, the squared sound pressure in the nth frequency band can be calculated from,

$$p_n^2 \approx \frac{W_n \rho c}{4\pi r^2} \times 10^{A_n(\tau,h)r/10\,000}, \tag{14}$$

where W_n denotes the sound power in the nth frequency band, ρc expresses the characteristic impedance of air, and r is the source-observer distance expressed in meters. Here $A_n(\tau,h)$ is the standardized attenuation coefficient at 1 atm pressure, depending on air temperature, τ, and relative humidity, h.[23] Formula (14) holds when the noise energy propagates along straight lines in the homogeneous atmosphere being at rest. Due to refraction the rays are curved and the right-hand side of this formula can be treated as a ''first approximation'' within 1 dB accuracy.[24] The A-weighted squared sound pressure is [Eq. (14)]

$$p_A^2 \approx \frac{W_A \rho c}{4\pi r^2} F_A(r,\tau,h), \tag{15}$$

where W_A expresses the A-weighted sound power and

$$F_A(r) = \sum_n 10^{\delta L_n/10} \times 10^{A_n r/10\,000}, \tag{16}$$

defines the A-weighted absorption factor. In Eq. (16),

$$\delta L_n = L_{Wn} - L_{WA} + \Delta L_n, \tag{17}$$

where L_{Wn} is the sound power level in the nth frequency band, and L_{WA} is the overall A-weighted power level. The relative power spectrum, $L_{Wn} - L_{WA}$, typical of aircraft noise, one can find in Ref. 25. The A-weightings are $\Delta L_n = -26.2$, -16.1, -8.6, -3.2, 0, $+1.2$, $+1.0$, and, -1.1 dB, for the octave bands that are characterized by their center frequencies, $f_n = 63,125,...,8000$ Hz, respectively. A prior knowledge of δL_n [Eq. (17)] and the standardized attenuation coefficients, A_n,[23] determines fully the function $F_A(r)$ [Eq. (16)]. Now, for a specified A-weighted sound level threshold, $\tilde{L}_A$, one obtains a transcendental equation in respect to distance, $\tilde{r}$ [Eq. (15)],

$$\frac{F_A(\tilde{r})}{\tilde{r}^2} = \frac{4\pi}{S_0} 10^{(\tilde{L}_A - L_{WA})/10}, \tag{18}$$

where $S_0 = 1 m^2$. Solution to Eq. (18) yields the radius, $\tilde{r}$, of the sphere which consists of points with $\tilde{L}_A = L_A$ (Fig. 5). At these points an aircraft noise begins or ceases to be annoyed. Cutting of the sphere, $r = \tilde{r}$, by the ground surface, $z = 0$, gives the circular threshold area in the horizontal plane $x-y$, with the radius,

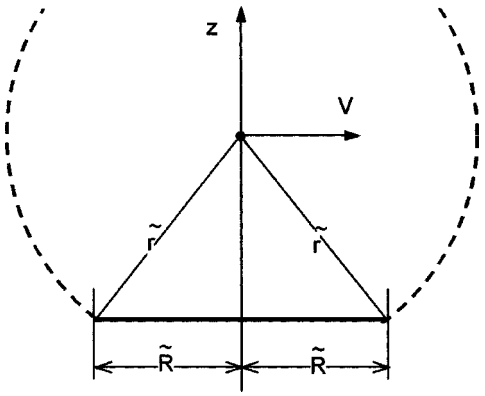


FIG. 5. The sphere of radius $\tilde{r}$ [Eq. (18)] and the threshold area of radius $\tilde{R}$ [Eq. (19)].

$$\tilde{R} = \sqrt{(\tilde{r})^2 - H^2}. \tag{19}$$

As an example, let us calculate $\tilde{r}$ for the relative power spectrum from Ref. 25. Taking the overall A-weighted power level, $L_{WA} = 140$ dB, the A-weighted sound level threshold, $\tilde{L}_A = 50$ dB, air temperature, $\tau = 0$ °C, and relative humidity, $h = 70\%$, one obtains, $\tilde{r} \approx 3500$ m [Eqs. (16)–(18)].

V. NOISE DURATION

The annoyance area (shaded region in Fig. 6) is common to the circular threshold area of radius $\tilde{R}$ [Eq. (19)] and the impact area which boundary is described by the function $R(\Phi - \Psi)$ [Eq. (12)]. Within the annoyance area the aircraft noise interferes with communication, natural quiet, and sound of nature. For an aircraft flying with the steady speed V, an observer is exposed to noise during the time interval, $\tau = L/V$. Here L is the length of the chord (parallel to the flight track) that connects two points of the boundary of annoyance area. For any wind direction with $0 < \Psi \leqslant 2\pi$ (Fig. 3), L can be found numerically (Fig. 6). Now, special cases, with closed form solutions, will be considered.

A. Windless atmosphere

When the atmosphere is at rest, $u = 0$, $\beta = 0$ [Eqs. (3) and (7)], the impact area is contained within the circle of

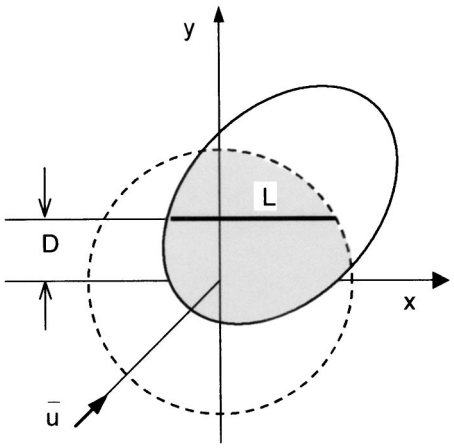


FIG. 6. The annoyance area (shaded region) is a common sector of the impact area [Eq. (12)] and the circular threshold area [Eqs. (18) and (19)].

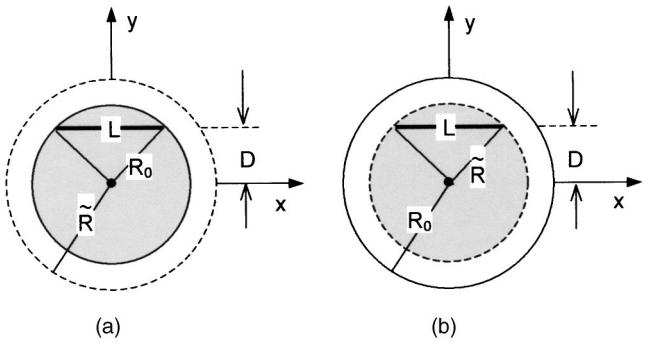


FIG. 7. The annoyance area equal to the impact area [Eq. (13)], $R_0<\tilde{R}$ (a), and to the threshold area [Eqs. (18), (19)], $\tilde{R}<R_0$ (b). Windless atmosphere, $\beta=0$.

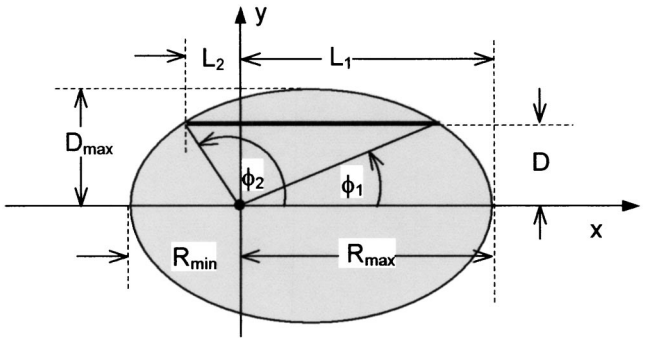


FIG. 9. Aircraft noise is annoyed within the impact area, $R(\Phi)$ [Eq. (22)]. Downwind flight with $\Psi=0$ (Fig. 3), weak wind, $\beta<\alpha$.

radius, R_0 [Eq. (13)]. If the radius $\tilde{R}$ [Fig. 5, Eq. (19)] is longer than R_0, then the impact area is a subset of the threshold area and the overflight time is [Fig. 7(a)],

$$\tau=\frac{2}{V}\sqrt{R_0^2-D^2}. \tag{20}$$

When the distance between the observer and the ground flight track meets the condition, $D>R_0=\sqrt{2H/\alpha}$, the aircraft is not heard at all. If the radius of the impact area, R_0, exceeds the radius of the annoyance area, $\tilde{R}$ [Fig. 7(b)], then the overflight time is given by

$$\tau=\frac{2}{V}\sqrt{\tilde{R}^2-D^2}. \tag{21}$$

B. Down- or upwind flight

Now suppose the wind blows along the flight path (the x axis) with $\Psi=0$ (Fig. 3). Hence Eq. (12) yields the boundary of the impact area,

$$R(\Phi)=R_0\left[1-\frac{\beta}{\alpha}\cos\Phi\right]^{-1/2}. \tag{22}$$

Two cases will be examined: closed (weak wind, $\beta<\alpha$) and open boundary of the impact area (strong wind, $\beta>\alpha$).

1. Weak wind, $\beta<\alpha$

When the temperature gradient overrides the wind speed gradient, $\beta<\alpha$, the shadow zone completely encircles the source and the impact area is contained within a closed curve (Fig. 8). The horizontal distance between the aircraft and the boundary of the impact area, $R(\Phi)$ [Eq. (22)], achieves its minimum and maximum,

$$R_{\min}=\frac{R_0}{\sqrt{1+\beta/\alpha}},\quad R_{\max}=\frac{R_0}{\sqrt{1-\beta/\alpha}}. \tag{23}$$

For $\tilde{R}<R_{\min}$, the threshold area is overlapped completely by the impact area [Fig. 8(a)]. In other words, the former is the subset of the latter and the overflight time can be calculated from Eq. (21). When $R_{\max}<\tilde{R}$, the impact area is a subset the threshold area [Fig. 8(b)]. In such a case the overflight time is (Fig. 9)

$$\tau=\frac{L_1+L_2}{V}, \tag{24}$$

with

$$L_1=D\operatorname{ctg}\Phi_1,\quad L_2=-D\operatorname{ctg}\Phi_2, \tag{25}$$

where the angles Φ_1 and Φ_2 are solutions to

$$R(\Phi)\sin\Phi=D. \tag{26}$$

Combining with Eq. (22) yields

$$\cos^2\Phi-\frac{\beta}{\alpha}\left(\frac{D}{R_0}\right)^2\cos\Phi+\left(\frac{D}{R_0}\right)^2-1=0. \tag{27}$$

Finally (Eq. 25),

$$L_1=R_0K_1\left[1-\frac{\beta}{\alpha}K_1\right]^{-1/2},\quad L_2=R_0K_2\left[1+\frac{\beta}{\alpha}K_2\right]^{-1/2}, \tag{28}$$

where

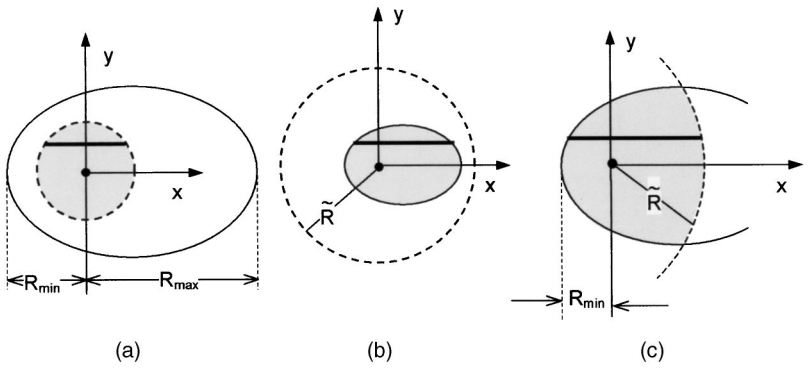


FIG. 8. The annoyance area is equal to the threshold area [(a) with $\tilde{R}<R_{\min}$—Eqs. (18), (19), (23)], to the impact area [(b) with $R_{\max}<\tilde{R}$—Eq. (23)], and to the common sector of the impact and threshold area [(c) with $R_{\min}<\tilde{R}<R_{\max}$]. Downwind flight with $\Psi=0$ (Fig. 3), weak wind, $\beta<\alpha$.

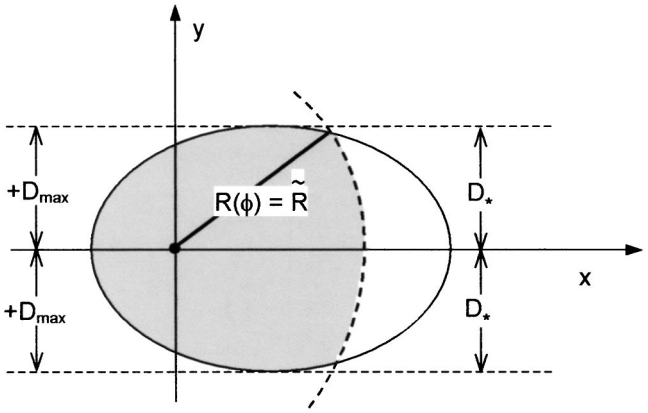


FIG. 10. The annoyance area (shaded region) is a common sector of the impact area [Eq. (12)] and the circular threshold area [Eqs. (18) and (19)]. Downwind flight with $\Psi=0$ (Fig. 3), weak wind, $\beta<\alpha$.

$$K_1=\sqrt{1-\frac{D^2}{R_0^2}+\left(\frac{\beta}{2\alpha}\frac{D^2}{R_0^2}\right)^2}+\frac{\beta}{2\alpha}\frac{D^2}{R_0^2},$$
$$K_2=\sqrt{1-\frac{D^2}{R_0^2}+\left(\frac{\beta}{2\alpha}\frac{D^2}{R_0^2}\right)^2}-\frac{\beta}{2\alpha}\frac{D^2}{R_0^2}. \tag{29}$$

Equations (24), (28), and (29) give an identical value of τ on each side of the ground flight track (the x axis), either $+D$ or $-D$, and the maximal value of τ for an overhead flight with $D=0$ (Fig. 9),

$$\tau_{\max}=\frac{R_0}{V}\left\{\frac{1}{\sqrt{1-\beta/\alpha}}+\frac{1}{\sqrt{1+\beta/\alpha}}\right\}. \tag{30}$$

In view of Eq. (23) the above-given formula can be rewritten as $\tau_{\max}=(R_{\max}+R_{\min})/V$. Aircraft noise is not heard when the perpendicular distance between the observer and the x axis exceeds (Fig. 9)

$$D_{\max}=\sqrt{2}R_0\frac{\alpha}{\beta}\sqrt{1-\sqrt{1-\frac{\beta^2}{\alpha^2}}},\quad \beta<\alpha. \tag{31}$$

At a perpendicular distance, $D=D_{\max}$, one gets $\tau=0$ [Eqs. (24), (28), (29), (31)]. When the radius of threshold area meets the conditions, $R_{\min}<\tilde{R}<R_{\max}$ [Fig. 8(c)],

$$\tau=\frac{1}{V}[\sqrt{\tilde{R}^2-D^2}+L_2(d)], \tag{32}$$

where $-D_*<D<+D_*$, the length L_2 is defined by Eqs. (28) and (29) and (Fig. 10),

$$D_*=\tilde{R}\sqrt{1-\frac{\alpha^2}{\beta^2}\left(1-\frac{R_0^2}{\tilde{R}^2}\right)^2},\quad \beta<\alpha. \tag{33}$$

Here R_0 and $\tilde{R}$ are defined by Eqs. (13) and (19), respectively. When the perpendicular distance between the listener and aircraft ground track satisfies the conditions, either $D_*<D<D_{\max}$, or, $-D_{\max}<D<-D_*$ (Fig. 10), one can calculate τ from Eqs. (24), (28), and (29).

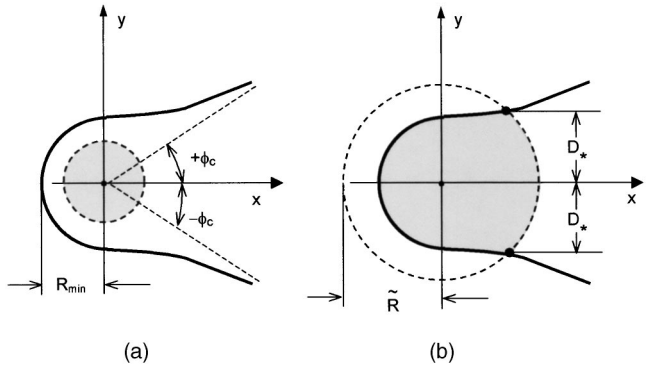


FIG. 11. The annoyance area is equal to the threshold area [(a) with $\tilde{R}<R_{\min}$—Eqs. (18), (19), (23)] and to the common sector of the impact and threshold area [(b) with $R_{\min}<\tilde{R}$]. Downwind flight with $\Psi=0$ (Fig. 3), strong wind, $\beta>\alpha$.

2. Strong wind, $\beta>\alpha$

When $\beta>\alpha$, the boundary of the impact area $R(\Phi)$ (Eq. 22) is open. For rays characterized by angles Φ within the range

$$-\Phi_c<\Phi<+\Phi_c, \tag{34}$$

with $\Phi_c=\cos^{-1}(\alpha/\beta)$,[22,26] the effect of wind outweighs that of the temperature gradient and no shadow zone occurs. In Fig. 11(a) the radius of the threshold area, $\tilde{R}$ (Eq. 19), is less than $R_{\min}$ (Eq. 23). Hence, due to complete overlapping, the aircraft noise is annoyed within the circle of radius $\tilde{R}$ and the overflight time, τ, is defined by Eq. (21). For $R_{\min}<\tilde{R}$ [Fig. 11(b)] partial overlapping occurs and one can calculate τ from Eq. (32). Aircraft noise is not heard at all, when $D>D_*$ and $D<-D_*$ [Eq. (33) with $\beta>\alpha$].

C. Crosswind flight

For wind blowing perpendicular to the flight path (along the y axis), with $\Psi=\pi/2$ shown in Fig. 3, Eq. (12) defines the boundary of the impact area,

$$R(\Phi)=R_0\left[1-\frac{\beta}{\alpha}\sin\Phi\right]^{-1/2}. \tag{35}$$

Similar to the case of down- or upwind flight, two cases will be considered: closed (weak wind, $\beta<\alpha$) and open boundary of the impact area (strong wind, $\beta>\alpha$).

1. Weak wind, $\beta<\alpha$

It is already known (Sec. V B) that inequality $\beta<\alpha$ brings about the impact area contained within a closed curve [Eq. (35)]. There is minimum, $R_{\min}$, and maximum horizontal distance, $R_{\max}$, between the aircraft ground track and the shadow zone [Eq. (23)], which characterize the geometry of an impact area. Taking into account relations between the radius of the threshold area, $\tilde{R}$, and the distances, $R_{\min}$, $R_{\max}$, there are three cases.

When $\tilde{R}<R_{\min}$ [Fig. 12(a)], the overflight time can be calculated from Eq. (21), because the impact area contains the whole threshold area. Similarly, for $R_{\max}<\tilde{R}$ the whole of impact area is located within the threshold area [Fig. 12(b)] and the overflight time can be calculated from

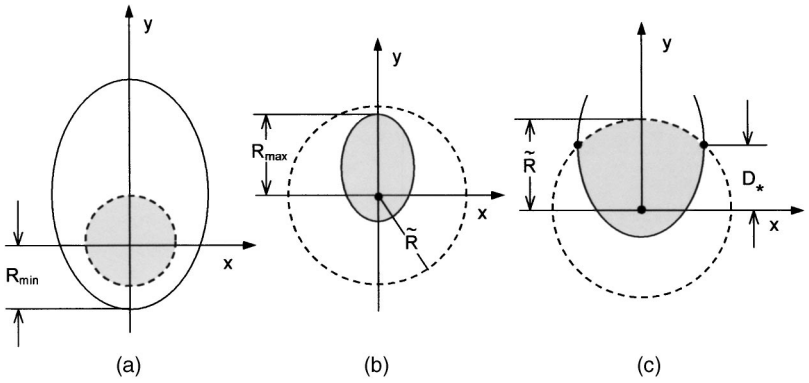


FIG. 12. The annoyance area is equal to the threshold area [(a) with $\tilde{R}<R_{\min}$—Eqs. (18), (19), (23)], to the impact area [(b) with $R_{\max}<\tilde{R}$—Eq. (23)], and to the common sector of the impact and threshold area [(c) with $R_{\min}<\tilde{R}<R_{\max}$]. Crosswind flight with $\Psi=\pi/2$ (Fig. 3), weak wind, $\beta<\alpha$.

$$\tau=\frac{2D}{V}\operatorname{ctg}\Phi, \tag{36}$$

where the Φ is the solution to Eq. (26) with $R(\Phi)$ defined by Eq. (35),

$$\sin^2\Phi+\frac{\beta}{\alpha}\left(\frac{D}{R_0}\right)^2\sin\Phi-\left(\frac{D}{R_0}\right)^2=0. \tag{37}$$

For the perpendicular distance, $\pm D$, the overflight time is (Fig. 13)

$$\tau_-=\frac{2}{V}\frac{\sqrt{R_0^2-D^2(\sqrt{1+q^2}+q)^2}}{\sqrt{1+q^2}+q},\quad -R_{\min}<D<0, \tag{38}$$

and

$$\tau_+=\frac{2}{V}\frac{\sqrt{R_0^2-D^2(\sqrt{1+q^2}-q)^2}}{\sqrt{1+q^2}-q},\quad 0<D<+R_{\max}, \tag{39}$$

where $q=(\beta/2\alpha)(D/R_0)$.

Either above or below the x axis, the aircraft noise disappears, $\tau=0$, when $D<-R_{\min}$ and $D>R_{\max}$ [Eq. (23)], respectively. An overhead flight with $D=0$ yields [Eqs. (38) and (39)]

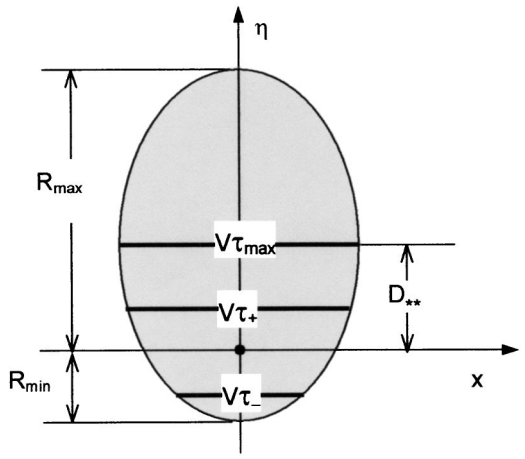


FIG. 13. Aircraft noise is annoyed within the impact area, $R(\Phi)$ [Eq. (35)]. Crosswind flight with $\Psi=\pi/2$ (Fig. 3), weak wind, $\beta<\alpha$.

$$\tau_+=\tau_-=\frac{2R_0}{V}. \tag{40}$$

However, contrary to the case for down- or upwind flight, the maximal overflight time does not occur for the flight above the x axis with $D=0$. For a crosswind flight under consideration,

$$\tau_{\max}=2\sqrt{2}\,\frac{R_0}{V}\,\frac{\alpha}{\beta}\sqrt{1-\sqrt{1-\frac{\beta^2}{\alpha^2}}}, \tag{41}$$

when the distance between the ground flight track and the observer equals

$$D_{**}=R_0\frac{\alpha}{\beta}\cdot\frac{1-\sqrt{1-\beta^2/\alpha^2}}{(1-\beta^2/\alpha^2)^{1/4}}. \tag{42}$$

When the strength of both gradients is the same, $\alpha=\beta$, the maximal duration, $\tau_{\max}$, is $\sqrt{2}$ longer than the noise duration for an overhead flight [Eqs. (40) and (41)]. In a windless atmosphere with $\beta=0$, Eq. (41) yields as expected, $\tau_{\max}=2R_0/V$, which means that the impact area takes a circular shape (Fig. 7). For $R_{\min}<\tilde{R}<R_{\max}$, there are two cases with the distance [Fig. 12(c), Eq. (35)]

$$D^*=\frac{\alpha}{\beta}\tilde{R}\left[1-\left(\frac{R_0}{\tilde{R}}\right)^2\right]. \tag{43}$$

When $-R_{\min}<D<D^*$, the overflight duration, τ, is given by Eqs. (38) and (39). For $D^*<D<\tilde{R}$, the time interval, τ, can be calculated from Eq. (21).

2. Strong wind, $\beta>\alpha$

When $\beta>\alpha$, the boundary of the impact area $R(\Phi)$ [Eq. (35)] is open. In Fig. 14(a) the radius of the threshold area, $\tilde{R}$ [Eq. (19)], is less than $R_{\min}$ [Eq. (23)] and the overflight time, τ, is defined by Eq. (21). For $R_{\min}<\tilde{R}$ [Fig. 14(b)] partial overlapping occurs. For $-R_{\min}<D<D^*$, Eqs. (38) and (39) can be employed for τ calculation. If $D^*<D<R_{\max}$, then Eq. (21) can be used again.

VI. CONCLUSIONS

It seems the percentage of time that an aircraft is audible, P [Eq. (1)], can be used to supplement the Day–Night Average Sound Level and the overflight time, τ, to supple-

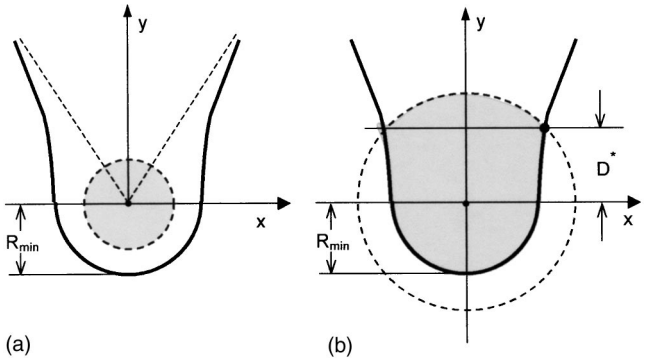


FIG. 14. The annoyance area is equal to the threshold area [(a) with $\tilde{R} < R_{\min}$—Eqs. (18), (19), (23)] and to the common sector of the impact and threshold area [(b) with $R_{\min} < \tilde{R}$]. Crosswind flight with $\Psi = \pi/2$ (Fig. 3), strong wind, $\beta > \alpha$.

ment the sound exposure level. For any wind direction (Fig. 3), τ can be found through numerical methods (Fig. 6). Closed form solutions are possible for a windless atmosphere, and down- and crosswind flights (Sec. V). The values of τ could provide guidance in the design of flight paths to minimize visitors judgment of annoyance. The paper ignores the audibility overlap of successive aircraft passbys. The second limitation of the paper comes from the assumption of the flat ground: sometimes the line-of sight between the aircraft and observer is interrupted (e.g., deep canyon or high hill).

[1] S. Fidell, L. Silvati, R. Howe, K.S. Pearsons, B. Tabachnik, R.C. Knopf, J. Gramann, and T. Buchanan, ''Effects of aircraft overflights on wilderness recreationists,'' J. Acoust. Soc. Am. **100**, 2909–2918 (1996).

[2] N.P. Miller, ''The effects of aircraft overflights on visitors to U.S. National Parks,'' Noise Control Eng. J. **47**, 112–117 (1999).

[3] J.B. Stokes, D.J. Leese, and S.L. Montgomery, ''Citizens get relief from recreation noise: The case in the skies from Hawaii,'' Noise Control Eng. J. **47**, 142–146 (1999).

[4] B. Berglund, K. Harder, and A. Preis, ''Annoyance perception of sound and information extraction,'' J. Acoust. Soc. Am. **95**, 1501–1509 (1994).

[5] G.A. Daigle, ''Acoustics of noise control outdoors,'' International Congress on Acoustics, Trondheim, Norway, 1995, pp. 49–56.

[6] K.E. Gilbert, R. Raspet, and X. Di, ''Calculation of turbulence effects on an upward-refracting atmosphere,'' J. Acoust. Soc. Am. **87**, 2428–2437 (1990).

[7] W.E. McBride, ''Scattering of sound by atmospheric turbulence: Predictions in the refractive shadow zone,'' J. Acoust. Soc. Am. **91**, 1336–1340 (1992).

[8] D.I. Havelock, ''Spatial coherence of sound field in a refractive shadow: Comparison of simulation and experiment,'' J. Acoust. Soc. Am. **98**, 2289–2302 (1995).

[9] K.M. Li, V.E. Ostashev, and K. Attenborough, ''The diffraction of sound in a stratified moving atmosphere above an impedance plane,'' Acust. Acta Acust. **84**, 607–615 (1998).

[10] M. West, J.D. Turton, and I.D. Todd, ''On the accuracy of ray tube model predictions for realistic meteorological conditions,'' Appl. Acoust. **42**, 277–290 (1994), Fig. 4.

[11] V.E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN SPON, London 1997), Fig. 1.4.a.

[12] A.D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).

[13] T.F.W. Embleton, G.J. Thiessen, and J.E. Piercy, ''Propagation in an inversion and reflection from the ground,'' J. Acoust. Soc. Am. **59**, 278–282 (1976).

[14] T. Hidaka, K. Kageyama, and S. Masuda, ''Sound propagation in the rest atmosphere with linear sound velocity profile,'' J. Acoust. Soc. Jpn. (E) **6**, 117–125 (1985).

[15] K.B. Rasmussen, ''Outdoor sound propagation under the influence of wind and temperature gradient,'' J. Sound Vib. **104**, 321–335 (1986).

[16] C.G. Don and A.J. Cramond, ''Comparison of ray and wave approaches to acoustics impulse propagation prior to a shadow boundary,'' J. Acoust. Soc. Am. **87**, 1017–1025 (1990).

[17] K.M. Li, K. Attenborough, and N.W. Heap, ''Source height determination by ground effect inversion in the presence of a sound velocity gradient,'' J. Sound. Vib. **145**, 111–128 (1991).

[18] A. L'Esperance, P. Herzog, G.A. Daigle, and J.R. Nicolas, ''Heuristic model for outdoor sound propagation based on the geometrical ray theory in the case of a linear sound profile,'' Appl. Acoust. **37**, 1403–1428 (1991).

[19] R. Raspet, A. L'Esperance, and G.A. Daigle, ''The effect of realistic ground impedance on the accuracy of ray tracing,'' J. Acoust. Soc. Am. **97**, 154–158 (1995).

[20] S. Taherzadeh, K.M. Li, and K. Attenborough, ''Some practical consideration for predicting outdoor sound propagation in the presence of wind and temperature gradients,'' Appl. Acoust. **54**, 27–44 (1998).

[21] F.M. Wiener and D.N. Keast, ''Experimental study of the propagation of sound over ground,'' J. Acoust. Soc. Am. **31**, 724–733 (1959).

[22] R. Makarewicz, ''The shadow zone in stratified media,'' J. Acoust. Soc. Am. **85**, 1092–1096 (1989).

[23] ANSI S1-1995 (ASA 113) American National Standard, ''Method for calculation of the absorption of sound by the atmosphere.''

[24] A.P. Dowling and J.E. Ffowcs Williams, *Sound and Sources of Sound* (Ellis Horwood, New York, 1983), pp. 119–121.

[25] G.S. Anderson and U.J. Kurze, ''Outdoor sound propagation,'' in *Noise and Vibration Control*, edited by L.L. Beranek and I.L. Ver (Wiley, New York, 1992), p. 115.

[26] U. Ingard, ''A review of the influence of meteorological conditions on sound propagation,'' J. Acoust. Soc. Am. **25**, 405–411 (1953).

Synthetic aperture imaging using sources with finite aperture: Deconvolution of the spatial impulse response

Fredrik Lingvall,[a] Tomas Olofsson, and Tadeusz Stepinski
Signals and Systems Group, Department of Material Science, Uppsala University, Box 528, 751 20 Uppsala, Sweden

(Received 19 April 2002; revised 3 March 2003; accepted 19 March 2003)

A method for ultrasonic synthetic aperture imaging using finite-sized transducers is introduced that is based on a compact, linear, discrete model of the ultrasonic measurement system developed using matrix formalism. Using this model a time-domain algorithm for deconvolution of the transducer's spatial impulse responses (SIRs) is developed that is based on a minimum mean square error (MMSE) criterion. The algorithm takes the form of a spatiotemporal filter that compensates for the SIRs associated with a finite-sized transducer at every point of the processed image. A major advantage of the proposed method is that it can be used for any transducer, provided that its associated SIRs are known. This is in contrast to the synthetic aperture focusing technique (SAFT), which treats the transducer as a point source. The performance of the method is evaluated with simulations and experiments, performed in water using a linear phased array. The results obtained using the proposed method are compared to those obtained with a classical time-domain SAFT algorithm. For a finite aperture source, it is clearly shown that the resolution obtained using the proposed method is superior to that obtained using the SAFT algorithm. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1575746]

PACS numbers: 43.60.Gk, 43.60.Pt, 43.20.El [JCB]

I. INTRODUCTION

The finite size of ultrasonic transducers used in ultrasonic imaging introduces a distortion that limits the lateral resolution of conventional B-scan imaging. The considerable work that has been done to improve the image quality can be categorized in two slightly different approaches: ultrasonic image restoration and synthetic aperture imaging (SAI).

In the ultrasonic image restoration approach, image distortion is assumed to result from a convolution of the undistorted image with the transducer's point spread function (PSF).[1,2] Deconvolution can then be used to improve the image quality if the PSF can be calculated or estimated. Many authors proposed using the Wiener filter for this purpose.[1,3,4]

The most often used SAI method, the synthetic aperture focusing technique (SAFT), has both time-domain[5] and frequency-domain[6] implementations. In the frequency-domain, SAFT is a special solution of the inverse scattering problem.[7] In medical applications, requirements concerning frame rate have led to solutions where a combination of a conventional electronic array focusing and SAFT imaging is used.[8]

Recently, there has been considerable interest in multi-element synthetic aperture imaging applied to medical array systems.[9–12] These systems are defocused to emulate a single-element spatial response with high acoustic power. This technique involves treating the focus of the transducer as a virtual source for SAFT.[13] The virtual source, which is aimed to produce approximately spherical waves over a certain aperture angle, can be treated as a point source.[10,11]

In this paper a new method is proposed, which combines both of the above-presented approaches. The conventional time-domain SAFT is a special case of the method that utilizes a synthetic aperture for deconvolution of the source's spatial impulse responses (SIRs). The method can be used with any finite transducer provided that the associated SIRs at the image points of interest are known. This is an advantage compared to the SAFT, which treats the transducer as a point source. The method proposed here is based on a discrete image formation model expressed in matrix form that is used for deriving an inverse spatiotemporal filter for processing ultrasonic B-scans. The resulting filter, which minimizes the mean squared reconstruction error of the imaging system, has a fixed structure for a given imaging configuration. Thus, the filter can be precomputed for each imaging configuration and the final processing does not require intensive computations. Spatiotemporal inverse filtering has also been proposed by Ebbini and Cain,[14] Tanter *et al.*,[15] and Aubry *et al.*[16] The inverse filters proposed by these authors performed regularized inversions in the frequency domain, using iterative and singular value decomposition techniques, to obtain the minimum-norm least-squares solution of the inverse problem. However, their solution does not consider the measurement noise. The inverse filter proposed here is computed entirely in the time domain and the regularization is derived from the properties of the measurement noise and the *a priori* knowledge of the scattering targets.

This paper is organized as follows: In Sec. II a short presentation is given of the conventional SAFT algorithm formulated in matrix form. A linear model for image formation expressed using sparse matrices is derived in Sec. III. This model is then used in Sec. IV to derive the inverse filter that minimizes the mean squared error of the imaging sys-

[a]Electronic mail: fl@signal.uu.se

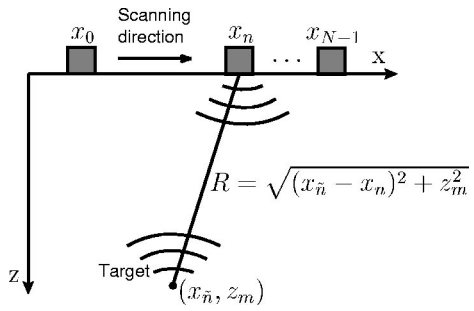


FIG. 1. Typical measurement setup for a SAFT experiment. The transducer is scanned linearly along the x axis and at each sampling position, $x_n = n\Delta x$, $n=0,1,...,N-1$, a data vector (A-scan) of M samples is recorded. The distance between the transducer, at $(x_n, z=0)$, and the target position $(x_{\tilde{n}}, z_m)$ is R.

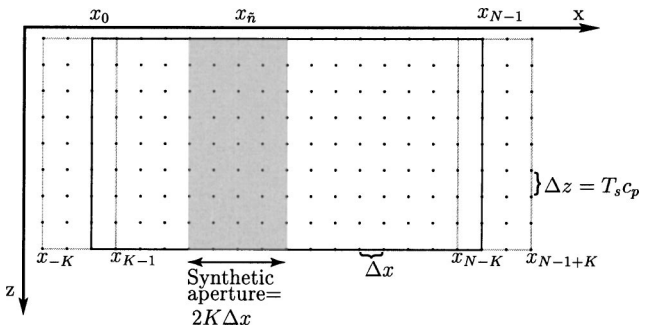


FIG. 2. Illustration of sampling. The ultrasonic B-scan data in $\mathbf{Y}$ are recorded for $x \in [x_0, x_{N-1}]$, shown as the solid-line rectangle. The reconstructed image, $\hat{\mathbf{O}}$, is calculated for every point. The shaded area shows the portion of $\mathbf{Y}$ that is used for reconstruction of the image points at $x_{\tilde{n}}$. The gray-line rectangles indicate the parts of $\hat{\mathbf{O}}$ for which the full synthetic aperture cannot be used. The vertical sampling distance Δz is determined by the temporal sampling period, T_s, and the sound speed, c_p, of the medium.

tem. The filter performance is illustrated by two experiments presented in Sec. V. First, a conventional SAFT and the proposed algorithm are compared using simulated results obtained for a strip line transducer. Then, a similar comparison is made for real measurements performed with an ultrasonic array imaging a wire target in water. Conclusions can be found in Sec. VI.

II. MATRIX FORMULATION OF THE SAFT ALGORITHM

Here, a matrix formulation of the time-domain SAFT algorithm is presented. This is done in the interest of clarity and SAFT is used to introduce the matrix tools used in the subsequent sections of this paper. Matrix formalism has been chosen here since it facilitates deriving the filter algorithm presented in Sec. IV. Furthermore, matrix formulation of SAFT enables a direct comparison with the proposed algorithm for better illustration of the relations between them.

The conventional time-domain SAFT algorithm performs synthetic focusing by means of coherent summations along hyperbolas (linear scanning of the transducer is assumed here). These hyperbolas simply express the distances, or time delays, from the transducer to the target positions (see Fig. 1).

Let Δx denote the spatial sampling interval along the scanning direction where an ultrasonic data vector (A-scan) of length M is recorded at each sampling position, $x_n = n\Delta x$, $n=0,1,...,N-1$, which gives an $M \times N$ measurement matrix $\mathbf{Y}$. Then, let $\hat{\mathbf{O}}$ denote the SAFT reconstructed image, where the caret over the variable denotes an estimate or a reconstruction. The reconstructed image takes the values at the spatial sampling points $x_{\tilde{n}} = \tilde{n}\Delta x$ and $z_m = m\Delta z$, $m = 0,...,M-1$, where $\Delta z = T_s c_p$, c_p is the sound speed of the medium, and T_s is the temporal sampling period (see Fig. 2).

Also, introduce the notation $(\mathbf{A})_{m,n}$ for element (m,n) in a matrix $\mathbf{A}$. The SAFT algorithm performs reconstruction by compensating for the propagation delay to each image point. This is achieved by a summation of $2K+1$ terms along hyperbolas according to (see Appendix A)

$$(\hat{\mathbf{O}})_{m,\tilde{n}} = \sum_{\substack{n=\tilde{n}-K \\ \{n|\tilde{n}-K\geq 0, \tilde{n}+K\leq N-1\}}}^{\tilde{n}+K} \alpha_n (\mathbf{Y})_{j,n}, \tag{1}$$

where α_n is an apodization weight[10] and j is the propagation delay index given by

$$j \doteq \frac{2}{c_p T_s}\sqrt{(x_{\tilde{n}}-x_n)^2+z_m^2}. \tag{2}$$

The operator $\doteq$ means that j is rounded toward the nearest integer and the synthetic aperture has a length $2K\Delta x$.

Now, let $\mathbf{y}_n = [u_o(x_n, t_0) u_o(x_n, t_1) \cdots u_o(x_n, t_{M-1})]^T$ denote the nth column vector in $\mathbf{Y}$ and define $d(\tilde{n},n)$ as the horizontal distance between $x_{\tilde{n}}$ and x_n, that is

$$d(\tilde{n},n) \triangleq x_{\tilde{n}} - x_n. \tag{3}$$

The delay-and-summation operation in Eq. (1) can now be expressed as a summation of matrix-vector multiplications; that is, the $\tilde{n}$th column, $\hat{\mathbf{o}}_{\tilde{n}}$, in $\hat{\mathbf{O}}$ can be expressed as

$$\hat{\mathbf{o}}_{\tilde{n}} = \sum_{\substack{n=\tilde{n}-K \\ \{n|\tilde{n}-K\geq 0, \tilde{n}+K\leq N-1\}}}^{\tilde{n}+K} \mathbf{S}_{d(\tilde{n},n)} \mathbf{y}_n, \tag{4}$$

where $\mathbf{S}_{d(\tilde{n},n)}$ is given by

$$(\mathbf{S}_{d(\tilde{n},n)})_{i,j} = \begin{cases} \alpha_{d(\tilde{n},n)} & \text{for } i = \dfrac{2z_m}{c_p T_s} \text{ and } j \text{ given by Eq. (2)} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The apodization weight $\alpha_{d(\tilde{n},n)}$ is normally a function of the distance $d(\tilde{n},n)$, or if a boxcar function is used, $\alpha_{d(\tilde{n},n)} \equiv 1$.

By defining $\mathbf{y} = \mathrm{vec}(\mathbf{Y}) \triangleq [\mathbf{y}_0^T \, \mathbf{y}_1^T \cdots \mathbf{y}_{N-1}^T]^T$, $\hat{\mathbf{o}} = \mathrm{vec}(\hat{\mathbf{O}})$, and noting that $d(\tilde{n},n) = d(\tilde{n}+i, n+i)$, for any integer i, the SAFT algorithm can finally be expressed as

$$\hat{\mathbf{o}} = \begin{bmatrix} \mathbf{S}_{d(K,0)} & & & \\ \vdots & \mathbf{S}_{d(K,0)} & \mathbf{0} & \\ \mathbf{S}_{d(-K,0)} & \vdots & & \\ & \mathbf{S}_{d(-K,0)} & \ddots & \\ & & \ddots & \mathbf{S}_{d(K,0)} \\ & \mathbf{0} & & \vdots \\ & & & \mathbf{S}_{d(-K,0)} \end{bmatrix} \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{N-1} \end{bmatrix}$$
$$= \mathbf{S}_{\mathrm{SAFT}}^{T}\mathbf{y}, \tag{6}$$

where the superscript T denotes the transpose. Note also that j is a function of $(x_{\tilde{n}} - x_n)^2$, implying the symmetry relation $\mathbf{S}_{d(-\tilde{n},0)} = \mathbf{S}_{d(\tilde{n},0)}$. The matrix $\mathbf{S}_{\mathrm{SAFT}}^{T}$ is an $M(2K+N)\times MN$ sparse block matrix.

The SAFT algorithm can be viewed as a matched filter,[17] where the filter associated with image point $(m,\tilde{n})$ is given by the hyperbolic function (2). The $\mathbf{S}_{\mathrm{SAFT}}$ matrix in Eq. (6) contains the matched filters for *every* image point. That is, by using Eq. (6) the matched filter outputs are computed simultaneously for all image points. Since the matched filter is the linear filter that maximizes the SNR for a given image point the SNR in the image after SAFT processing will increase.

A deficiency of the SAFT algorithm is that the transducer must be small compared to the wavelength. Otherwise, the filters in Eq. (6) will poorly match the true response of the imaging system. In many situations a finite-sized transducer must be used to obtain sufficient acoustic power. However, to obtain high resolution in this case the effects of the transducer size must be taken into account. Here, this is accomplished by formulating an estimation problem using a linear discrete model of the measuring system that takes into account the SIRs of the finite-sized transducer.

III. A LINEAR DISCRETE MODEL OF PULSE-ECHO MEASUREMENTS FOR HARD POINT TARGETS

In the following a discrete linear imaging model for finite-sized transducers is derived. This model will then be used for deriving the minimum mean square error solution for the reconstruction problem.

A. Diffraction model for a point-like target

The linear model of the imaging system is derived by studying a point target at position $\mathbf{r}_T$, with an elementary surface s_e, as depicted in Fig. 3.

The electrical input to the transducer is denoted $u_i(t)$ and its output $u_o(t)$. The transducer has a *forward* electro-acoustical impulse response denoted $h_{ef}(t)$ and a *backward* acousto-electrical impulse response $h_{eb}(t)$, respectively. By assuming a uniform piston transducer the normal velocity on its active area, $v(t)$, can be expressed as

$$v(t) = h_{ef}(t) * u_i(t), \tag{7}$$

where the asterisk denotes temporal convolution. In reception the transducer is sensitive to the total (instantaneous) pressure on its surface. The output $u_o(t)$ can be expressed as a convolution of $h_{eb}(t)$ with the total pressure, denoted $\langle p\rangle(\mathbf{r}_T,t)$, that is

$$u_o(t) = h_{eb}(t) * \langle p\rangle(\mathbf{r}_T,t). \tag{8}$$

The total pressure is the incident pressure, at time t, integrated over the transducer's active area, S_R.

To derive $\langle p\rangle(\mathbf{r}_T,t)$ an expression for the incident pressure at $\mathbf{r}_T$ must first be obtained. The pressure field at $\mathbf{r}_T$ can, using Huygen's principle, be expressed as[18]

$$p(\mathbf{r}_T,t) = \rho_0 v(t) * \frac{\partial}{\partial t}\left(\int_{S_R} \frac{\delta(t-(|\mathbf{r}_T-\mathbf{r}_R|/c_p))}{2\pi|\mathbf{r}_T-\mathbf{r}_R|} dS_R \right), \tag{9}$$

where $\mathbf{r}_R \in S_R$, and the total pressure received by the transducer from a point target at $\mathbf{r}_T$ can then be expressed as[18]

$$\langle p\rangle(\mathbf{r}_T,t) = \frac{s_e}{2c_p} p(\mathbf{r}_T,t) * \frac{\partial}{\partial t}\left(\int_{S_R} \frac{\delta(t-(|\mathbf{r}_R-\mathbf{r}_T|/c_p))}{2\pi|\mathbf{r}_R-\mathbf{r}_T|} dS_R \right). \tag{10}$$

Now, let $h(\mathbf{r},t)$ denote the (velocity) potential impulse response

$$h(\mathbf{r},t) \triangleq \int_{S_R} \frac{\delta(t-(|\mathbf{r}-\mathbf{r}_R|/c_p))}{2\pi|\mathbf{r}-\mathbf{r}_R|} dS_R.$$

Then, Eq. (10) can be written

$$\langle p\rangle(\mathbf{r}_T,t) = s_e \frac{\partial^2}{\partial t^2} h^{\mathrm{SIR}}(\mathbf{r}_T,t) * v(t), \tag{11}$$

where the double path *acoustical or spatial impulse response* (SIR), $h^{\mathrm{SIR}}(\mathbf{r}_T,t)$, is defined as

$$h^{\mathrm{SIR}}(\mathbf{r}_T,t) \triangleq \frac{1}{2\rho_0 c_p} h(\mathbf{r}_T,t) * h(\mathbf{r}_T,t). \tag{12}$$

Equation (11) describes both the transmission and reception diffraction process for scattering from point-like targets. Furthermore, define the *electrical impulse response*, $h_e(t)$, as

$$h_e(t) \triangleq \frac{\partial^2}{\partial t^2} h_{eb}(t) * h_{ef}(t). \tag{13}$$

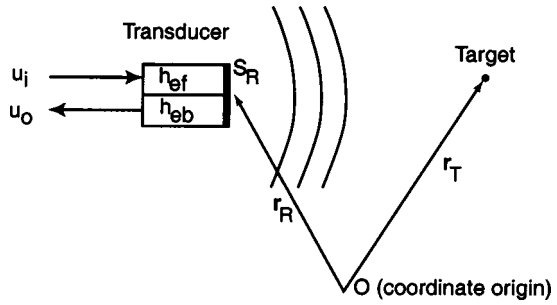


FIG. 3. Schematic view of the ultrasonic pulse-echo measurement system. The system has a forward electro-acoustical impulse response $h_{ef}(t)$, a backward acousto-electrical impulse response $h_{eb}(t)$, and the active area of the transducer is denoted S_R. The point target has an elementary surface s_e.

Then, by inserting Eq. (7) into Eq. (11) and Eq. (11) into Eq. (8), and by using Eqs. (12) and (13) the output, $u_o(t)$, can be expressed as

$$u_o(t) = s_e h^{SIR}(\mathbf{r}_T, t) * h_e(t) * u_i(t). \tag{14}$$

In conclusion, if the shape of the transducer surface, the electrical impulse response, and the positions of the targets are known, the (noise free) output $u_o(t)$ can be modeled using Eq. (14).

B. The discrete two-dimensional model

Here, the analysis is limited to the two-dimensional case, that is, the responses from targets outside the plane defined by $y=0$ are neglected (this can in practice be accomplished by using a suitable transducer with a very narrow field in the *y*-direction). Also, let the data collection be performed in the manner described in Sec. II. Then the signal received at point x can be expressed as

$$u_o(x,t) = s_e \sum_{(x_T, z_T) \in \mathcal{T}} h^{SIR}(x - x_T, z_T, t) * h_e(t) * u_i(t), \tag{15}$$

where $\mathcal{T}$ is the set of points, in the plane defined by $y=0$, where the targets are located. If multiple scattering is disregarded, then the model (15) takes both the electrical and acoustical properties into account when the scattering can be considered as a sum of contributions from individual point targets.

A time-discrete version of Eq. (15) is obtained by sampling the SIRs and the electrical impulse response, and replacing the continuous time convolutions in Eq. (15) by discrete time convolutions. A discrete representation of the targets is obtained by defining the so-called *object function* $o(x,z)$ as

$$o(x,z) \triangleq \begin{cases} s_e & \text{for } x,z \in \mathcal{T} \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

If $o(x,z)$ is discretized, it only takes values at the sampling points, hence, $o(x,z)$ can be represented by a matrix $\mathbf{O}$ with the same size as $\hat{\mathbf{O}}$ defined in Sec. II. Also, let the indices for the transducer position and image point be defined as in Sec. II. Note that several point targets can be within one resolution cell leading to the possibility of a varying target strength. That is, the nonzero elements in $\hat{\mathbf{O}}$ do not necessarily have the same value.

Now, consider the signal received from all point targets at the line $x = x_{\tilde{n}}$ and assume for the moment that $h_e(t) = u_i(t) = \delta(t)$ (recall that the targets on that line are represented by $\mathbf{o}_{\tilde{n}}$). The convolution and summation in Eq. (15) can then be expressed as a matrix-vector multiplication of the form (see Appendix B)

$$\begin{aligned} \mathbf{y}_n &= [\mathbf{h}_0^{SIR}(d(\tilde{n}, n)) \;\; \mathbf{h}_1^{SIR}(d(\tilde{n}, n)) \cdots \mathbf{h}_{M-1}^{SIR}(d(\tilde{n}, n))]\mathbf{o}_{\tilde{n}} \\ &= \mathbf{P}_{d(\tilde{n},n)} \mathbf{o}_{\tilde{n}}. \end{aligned} \tag{17}$$

where the column vector $\mathbf{h}_m^{SIR}(d(\tilde{n},n))$ contains the discretized spatial impulse response for depth z_m at the distance $d(\tilde{n},n)$. The $M \times M$ SIR matrix $\mathbf{P}_{d(\tilde{n},n)}$ then depends on the distance $d(\tilde{n},n)$ from the center of the transducer to $\mathbf{o}_{\tilde{n}}$.

Since the transducer will receive echoes from targets at more than one x position a summation must be performed over $\tilde{n}$. For practical transducers, the acoustic power decreases as the horizontal distance from the transducer increases; therefore, can the summation be truncated according to

$$\mathbf{y}_n = \sum_{\tilde{n}=n-L}^{n+L} \mathbf{P}_{d(\tilde{n},n)} \mathbf{o}_{\tilde{n}} \tag{18}$$

for a suitable choice of L without causing significant errors.

In reality the received signal will be band-limited due to the band-limited electrical impulse response $h_e(t)$. It will also be contaminated by measurement and quantization noise in the digitalization process. If the electrical impulse response is introduced into Eq. (18), and the noise is assumed to be additive, the A-scan measurement vector can be expressed as (see Appendix B)

$$\mathbf{y}_n = \mathbf{H} \sum_{\tilde{n}=n-L}^{n+L} \mathbf{P}_{d(\tilde{n},n)} \mathbf{o}_{\tilde{n}} + \mathbf{e}_n, \tag{19}$$

where $\mathbf{e}_n$ is the noise vector, and $\mathbf{H}$ is an $M \times M$ finite impulse response matrix defined by the *system* impulse response $h(t) \triangleq h_e(t) * u_i(t)$.

Equation (19) is a discrete time model for a single A-scan measurement. By using $\mathbf{o}$ and $\mathbf{y}$, defined in Eq. (6), the extension to a B-scan can easily be performed according to

$$\mathbf{y} = \mathbf{H}_0 \mathbf{S}_{SIR} \mathbf{o} + \mathbf{e}, \tag{20}$$

where $\mathbf{H}_0$ is a block diagonal $MN \times MN$ matrix with N $\mathbf{H}$:es along its diagonal. If the transducer is axisymmetric, then $\mathbf{S}_{SIR}$ has a similar structure to $\mathbf{S}_{SAFT}$ in Eq. (6) with $\mathbf{S}_{d(\tilde{n},n)}$ replaced by $\mathbf{P}_{d(\tilde{n},n)}$. Equation (20) is a discrete linear model for two-dimensional ultrasonic imaging when the backscattering is considered as a sum of responses from point targets. The model (20) can now be used to obtain simulated B-scan data by simply inserting an object function into Eq. (20) and rearranging $\mathbf{y}$ into an image. In the following, this model will be used for defining a filter that minimizes the reconstruction error.

IV. THE INVERSE FILTER

The SIRs and the system impulse response introduce a smearing which results in an unsharp ultrasonic image $\mathbf{Y}$. The method proposed here for improving the resolution consists in multiplying the measured data $\mathbf{y}$ with a reconstruction filter matrix $\mathbf{K}$ so that the reconstruction error after multiplication is minimized. More specifically, the matrix $\mathbf{K}$ is chosen so that the mean squared error $J = E\{\|\mathbf{o} - \mathbf{K}\mathbf{y}\|^2\}$ of the imaging system is minimized. The discretized object function $\mathbf{o}$ and the measurement noise $\mathbf{e}$ are assumed to be mutually independent zero mean stochastic processes with covariance matrices $\mathbf{C}_{oo}$ and $\mathbf{C}_{ee}$, respectively. Then by using Eq. (20) the mean squared error can be expressed as

$$J = E\{\|\mathbf{o} - \mathbf{K}\mathbf{y}\|^2\} = \mathrm{tr}\{\mathbf{C}_{oo}\} - 2\,\mathrm{tr}\{\mathbf{K}\mathbf{H}_0\mathbf{S}_{\mathrm{SIR}}\mathbf{C}_{oo}\} + \mathrm{tr}\{\mathbf{K}(\mathbf{H}_0\mathbf{S}_{\mathrm{SIR}}\mathbf{C}_{oo}\mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T + \mathbf{C}_{ee})\mathbf{K}^T\}, \tag{21}$$

where $\mathrm{tr}\{\cdot\}$ is the trace operator and $E\{\cdot\}$ is the expectation operator. The filter matrix, $\hat{\mathbf{K}}$, which minimizes Eq. (21), is the minimum mean square error estimator (MMSE) for the linear model (20),[19]

$$\hat{\mathbf{K}} = \arg\min_{\mathbf{K}} J = \mathbf{C}_{oo}\mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T(\mathbf{H}_0\mathbf{S}_{\mathrm{SIR}}\mathbf{C}_{oo}\mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T + \mathbf{C}_{ee})^{-1} = (\mathbf{C}_{oo}^{-1} + \mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T\mathbf{C}_{ee}^{-1}\mathbf{H}_0\mathbf{S}_{\mathrm{SIR}})^{-1}\mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T\mathbf{C}_{ee}^{-1}. \tag{22}$$

Assume, for a moment, that both **o** and **e** are white Gaussian processes, then their covariance matrices can be written as $\mathbf{C}_{ee} = \lambda_e\mathbf{I}$ and $\mathbf{C}_{oo} = \lambda_o\mathbf{I}$, respectively. The ratio between λ_o and the measurement noise variance, λ_e, can be seen as a tuning parameter which can be altered to trade between the confidence to the measurements and regularity of the solution. If, for example, the ratio is small, the diagonal matrix $\mathbf{C}_{ee}$ will dominate in the inverse in Eq. (22) resulting in smooth reconstructions. If the ratio is high, the inverse is dominated by the factor $\mathbf{H}_0\mathbf{S}\mathbf{C}_{oo}\mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T$, which means that more weight is given to the measurements.

Now, consider the following three special cases of Eq. (22).

(1) $\mathbf{H}_0\mathbf{S}_{\mathrm{SIR}}\mathbf{C}_{oo}\mathbf{S}^T\mathbf{H}_0^T$ is small compared to $\mathbf{C}_{ee}$ and $\mathbf{C}_{oo} = \mathbf{I}$. In this case Eq. (22) reduces approximately to $\hat{\mathbf{K}} = \mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T\mathbf{C}_{ee}^{-1}$, which is a matrix implementation of a matched filter.

(2) $\mathbf{H}_0 = \mathbf{I}$. If the system impulse response is disregarded $[h(t) = \delta(t)]$ the MMSE filter will only compensate for the SIR part of Eq. (20) and the filter can be viewed as a SAI algorithm that corrects diffraction effects.

(3) Combination of (1) and (2), $\mathbf{C}_{ee} = \lambda_e\mathbf{I}$, and a small sized transducer. When the transducer becomes small its associated discretized SIRs will approach $(s_e/\lambda_e R^2)^{-1}\delta(m-j)$, where R is defined in Fig. 1 and j is given by Eq. (2). As a result, the algorithm reduces to the delay-and-sum SAFT algorithm with an apodization correction for the rate of decrease of the spherical wave propagation. If this correction is applied to the SAFT algorithm (6) as well, then the algorithms become identical, that is, $\hat{\mathbf{K}} = \mathbf{S}_{\mathrm{SIR}}^T = \mathbf{S}_{\mathrm{SAFT}}^T$.

Thus, if the noise **e** is large the MMSE filter becomes cautious and reduces to the matched filter $\mathbf{S}_{\mathrm{SIR}}^T\mathbf{H}_0^T\mathbf{C}_{ee}^{-1}$ resulting in smooth reconstructions. Moreover, this matched filter reduces to the SAFT algorithm when the system impulse response approaches a Dirac function and the transducer can be regarded as a point source.

As mentioned in Sec. II, the matched filter maximizes the SNR for a given image point. The general form of a matched filter, for image point $(m,\tilde{n})$ for noise with zero mean and a covariance matrix **C**, is[20]

$$T_{m,\tilde{n}}(\mathbf{y}) = \mathbf{s}_{m,\tilde{n}}^T\mathbf{C}^{-1}\mathbf{y}, \tag{23}$$

where $\mathbf{s}_{m,\tilde{n}}$ is the so-called *template*. The scalar output $T_{m,\tilde{n}}(\mathbf{y})$, known as the *test statistic* in detection theory, maximizes the SNR for image point $(m,\tilde{n})$ without regarding other image points. Thus, for matched filters constraints such as main lobe width and sidelobe levels are not considered.[21]

It is, therefore, important to notice that there is a fundamental difference between maximizing the SNR at a fixed point and minimizing the *overall* reconstruction error[15,16]—the first approach leads to the spatiotemporal matched filter (STMF) solution and the latter to the MMSE solution. Since a matched filter only optimizes SNR at a single point[22,23] it will not necessarily result in high resolution.

The MMSE filter, on the other hand, is optimized to give the minimum reconstruction error, in the mean squared sense, for the imaging system at hand. That is, as long as the SIRs and the system impulse response are known the MMSE filter will consistently yield a solution with equal or better resolution than that of the STMF.

V. EXPERIMENTS

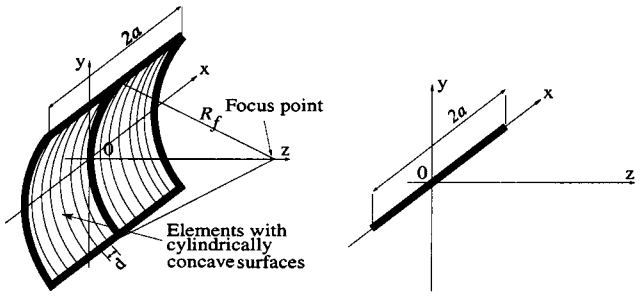
To illustrate the above-presented theoretical results two experiments have been performed. Simulated data were used in the first experiment and, in the second, real measurements; both experiments were aimed at imaging a single point target in water. The SAFT and the MMSE algorithm were compared using two different transducer sizes and two different focal depths. The first transducer size was 1 mm, which could be considered as a small-sized source and the second, 16 mm, is large enough to manifest significant beam directivity. The focal depths chosen were 50 and 190 mm, where the first depth is common in medical applications, and the second is common in, for example, NDT applications. The second depth, 190 mm, was also chosen since the transducer that was used in the measurements has a geometrical focus in the *y*-direction at 190 mm which minimizes the *y*-direction diffraction effects at that depth.

The simulated data were generated using Eq. (20) with $\mathbf{H}_0 = \mathbf{I}$ and $(\mathbf{O})_{m,\tilde{n}} = 1$ for $(x_{\tilde{n}} = 0,\ z_m = 50\,\mathrm{mm})$ and $(x_{\tilde{n}} = 0,\ z_m = 190\,\mathrm{mm})$, respectively, and zero otherwise. This simulates the (noise free) response from a single point target for an infinite bandwidth system, excited by a Dirac pulse.

The transducer used in the experiments, shown in Fig. 4(a), is a cylindrically concave phased array with a *y*-direction geometrical focus at 190 mm and a center frequency of 3 MHz. The use of an array facilitates the change of transducer size by simply selecting the number of active elements (all active elements were fired simultaneously), without changing the system impulse response.

The array element size d is approximately 1 mm and the maximum number of elements that can be used simultaneously is 32. Thus, by using the phased array an unfocused transducer in the $x-z$ plane could be imitated with a size of 1–32 mm (a thorough description of the array can be found in Ref. 24). The data were acquired using an ALLIN system which has an 8 bit analog-to-digital converter.

The measurements were performed in water, with $c_p = 1477\,\mathrm{m/s}$, and the target was a small steel pin (a wire



(a) Cylindrically concave transducer ($R_f = 190$ mm, $d = 1$ mm)

(b) Line strip transducer

FIG. 4. Transducer geometries.

target) with a diameter of 0.3 mm; the axis of the pin was parallel to the *y*-axis.

The spatial impulse responses, which were used both for the simulated and measured data, were calculated using an analytical solution for a uniformly exited line strip model.[25] The geometry of the line strip is shown in Fig. 4(b). The calculations were based on the assumption that the transducer acts as a plane piston that reduces to a line strip in two dimensions.

The double path spatial impulse response of a uniformly excited line strip is given by[25]

$$h_{\mathrm{c}}^{\mathrm{SIR}}(x,z,t)=g(x,t)\frac{2}{\sqrt{t^2-t_z^2}}*g(x,t)\frac{2}{\sqrt{t^2-t_z^2}}, \tag{24}$$

where the definition of the gating function, $g(x,t)$, is shown in Table I with $t_z=z/c_{\mathrm{p}}$.

A discrete version of Eq. (24) was obtained by summing all contributions from the SIR in the actual sampling interval. That is, the sampled SIR is defined as

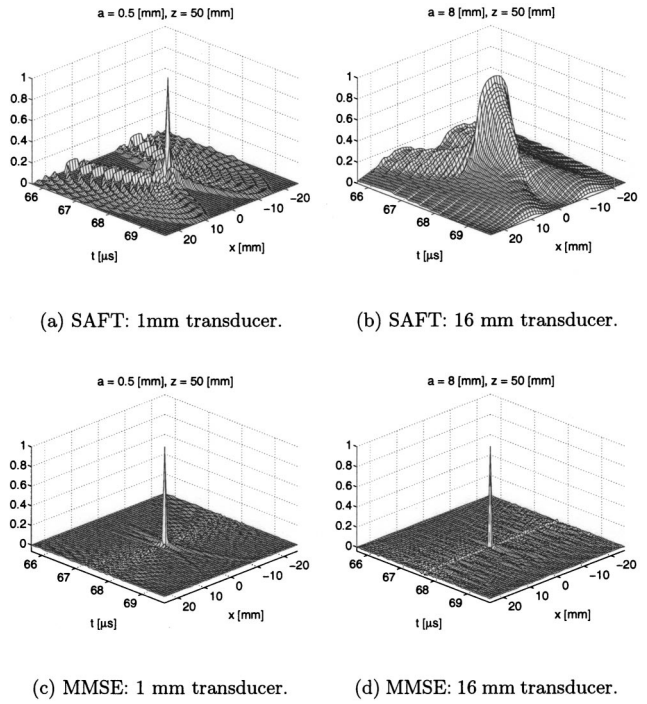
$$h^{\mathrm{SIR}}(x,z,t_m)\triangleq\int_{t_m}^{t_m+T_{\mathrm{s}}}h_{\mathrm{c}}^{\mathrm{SIR}}(x,z,t)dt. \tag{25}$$

Figures 5 and 6 show the results from the simulations while the results obtained for the measured data are shown in Figs. 7 and 8.

The results from the MMSE method are given without compensation for the system impulse response; that is, $\mathbf{H}_0$ in Eq. (22) was set to the identity matrix. This choice $\mathbf{H}_0$ was made to facilitate the comparison with the SAFT algorithm that does not provide any compensation for the system impulse response. The matrix $\mathbf{C}_{oo}$ in Eq. (22) was also chosen as $\mathbf{C}_{oo}=\mathbf{I}$, which is a conservative setting, and the noise was assumed white with variance λ_e. The parameter λ_e was used as a tuning parameter and it was adjusted manually to obtain the best visible performance ($\lambda_e=10^{-6}$ was used for the measurements and $\lambda_e=10^{-8}$ for the simulations). The system was sampled with $\Delta x=1$ mm and $T_{\mathrm{s}}=0.04\,\mu$s

TABLE I. Gating function for the spatial impulse response of a uniformly exited line strip.

$t_{1,2}=t_z\sqrt{1+(a\mp\lvert x\rvert/z)^2}$	
$\lvert x\rvert\leqslant a$	$\lvert x\rvert>a$
$g(x,t)=\begin{cases}2 & \text{for } t_z\leqslant t\leqslant t_1\\ 1 & \text{for } t_1<t\leqslant t_2\\ 0 & \text{otherwise}\end{cases}$	$g(x,t)=\begin{cases}1 & \text{for } t_1<t\leqslant t_2\\ 0 & \text{otherwise}\end{cases}$



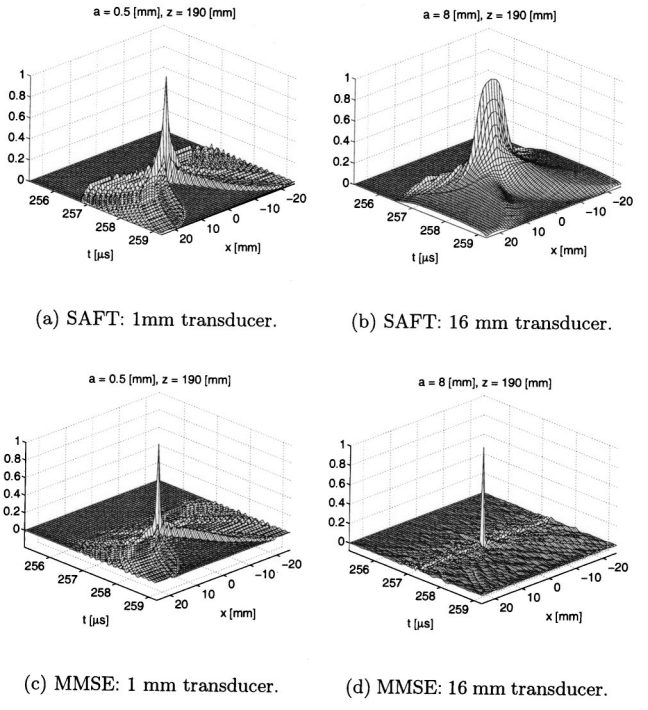
(a) SAFT: 1mm transducer.

(b) SAFT: 16 mm transducer.

(c) MMSE: 1 mm transducer.

(d) MMSE: 16 mm transducer.

FIG. 5. Comparison of the SAFT and the MMSE algorithm on simulated data from a point target at $z=50$ mm.



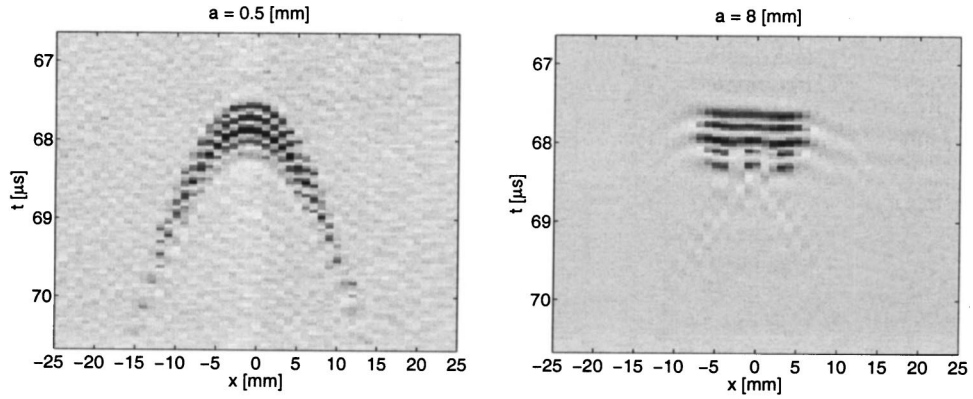
(a) SAFT: 1mm transducer.

(b) SAFT: 16 mm transducer.

(c) MMSE: 1 mm transducer.

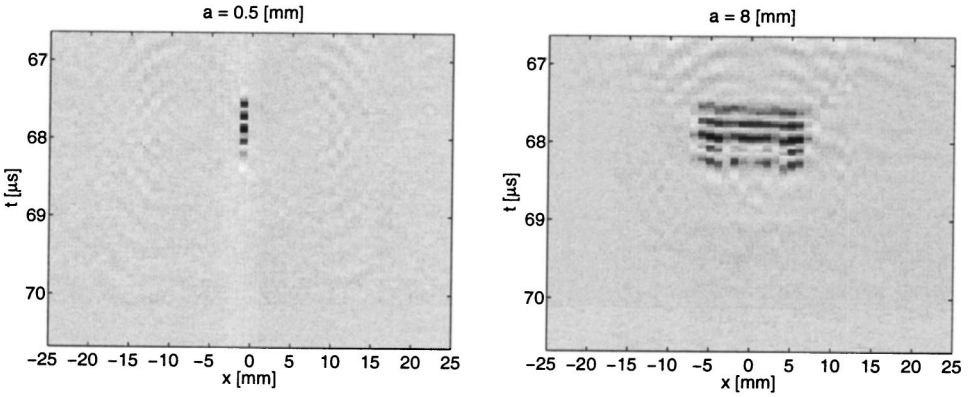
(d) MMSE: 16 mm transducer.

FIG. 6. Comparison of the SAFT and the MMSE algorithm on simulated data from a point target at $z=190$ mm.



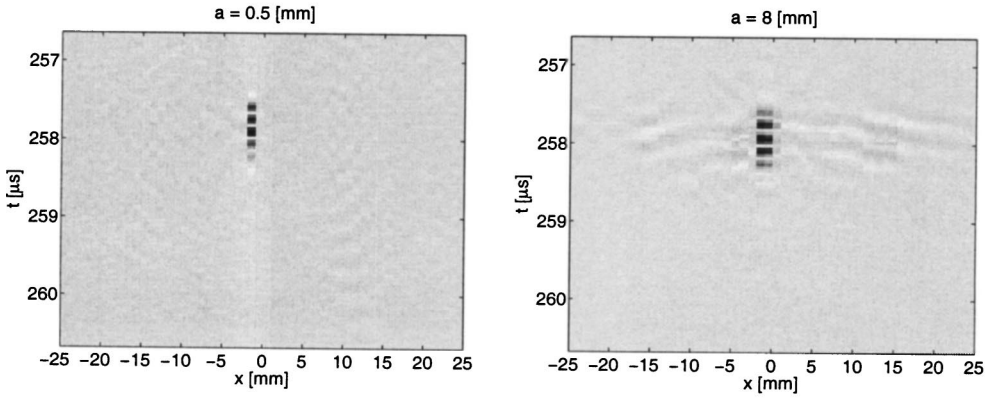
(a) Original: 1mm transducer

(b) Original: 16 mm transducer



(c) SAFT: 1mm transducer.

(d) SAFT: 16 mm transducer.



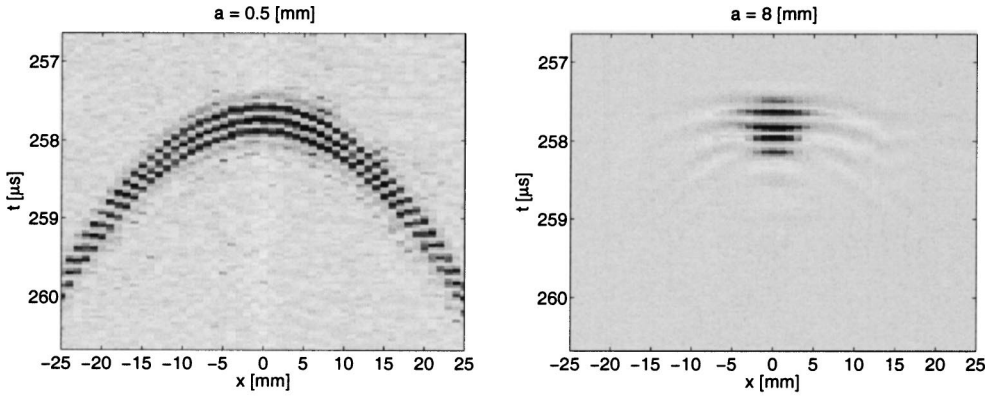
(e) MMSE: 1 mm transducer.

(f) MMSE: 16 mm transducer.

FIG. 7. Comparison of the SAFT and the MMSE algorithm on measured data from a point target (steel pin with a diameter of 0.3 mm) at $z=50$ mm using a cylindrically concave phased array.

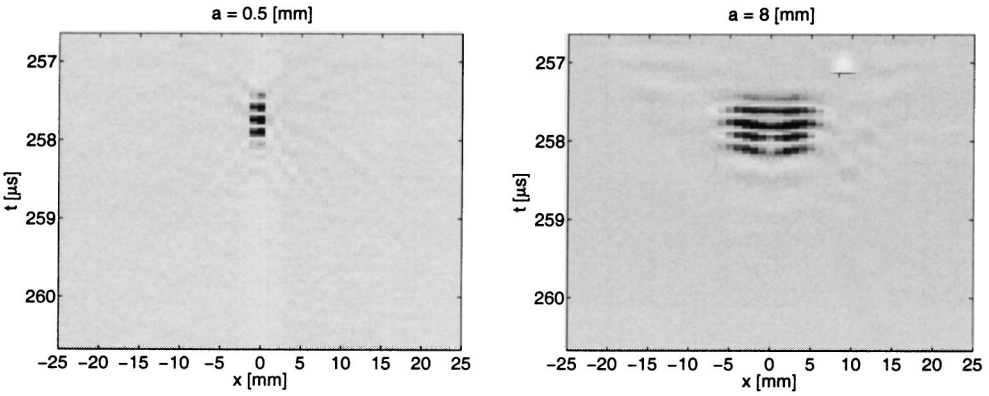
(=25 MHz) and 31 A-scans were used in the synthetic aperture for both the SAFT and the MMSE algorithm.

By comparing (a) and (c) in Figs. 5 and 6, which show the results from simulated data for the transducer of a 1 mm width, one can conclude that the SAFT algorithm has a similar resolution to the MMSE method. However, when the transducer size is increased to 16 mm (cf. (b) and (d) in Figs. 5 and 6) the SAFT produces overcompensated reconstructions without significant resolution improvement whilst the MMSE method performs almost as well as for the 1 mm transducer. The same behavior is also pronounced in the reconstructions from the measured data presented in Figs. 7 and 8. In these figures, (a) and (b) show the original measurements, (c) and (e) show the reconstructions from the 1



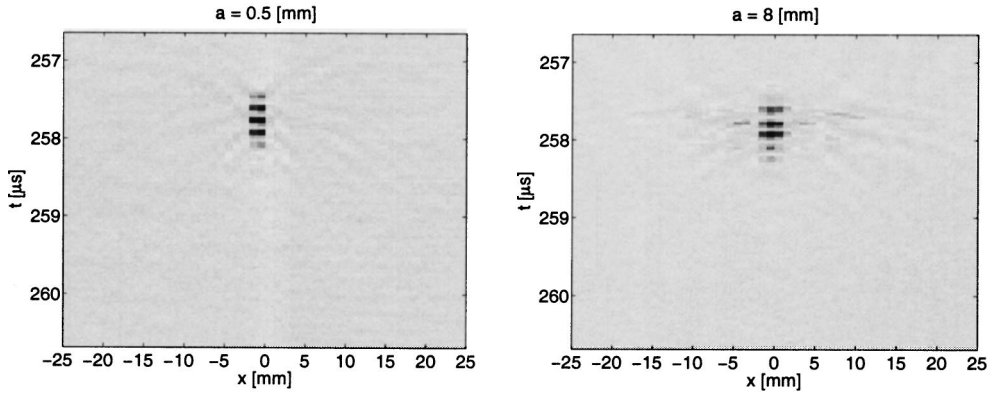
(a) Original: 1mm transducer

(b) Original: 16 mm transducer



(c) SAFT: 1mm transducer.

(d) SAFT: 16 mm transducer.



(e) MMSE: 1 mm transducer.

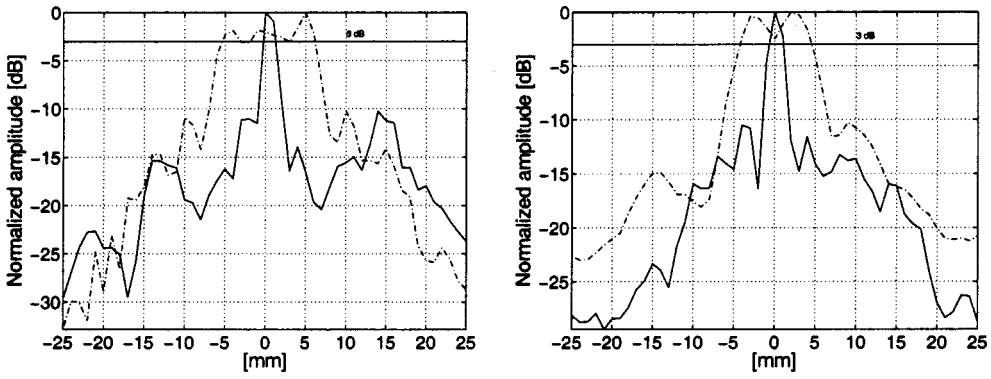
(f) MMSE: 16 mm transducer.

FIG. 8. Comparison of the SAFT and the MMSE algorithm on measured data from a point target (steel pin with a diameter of 0.3 mm) at $z = 190$ mm using a cylindrically concave phased array.

mm transducer, and (d) and (f) show the reconstructions from the 16 mm transducer. This is even more clearly seen in the profile plots in Fig. 9 where the 3 dB width is three to four times larger for the SAFT compared to that of the MMSE algorithm (for the 16 mm transducer).

It is worth noting that the SIRs that have been used here have been obtained solely using the theoretical model (25).

The equation systems (22) that must be solved in order to use the MMSE method are inherently large. The MMSE solution has, however, two properties that facilitate the computation; first, the matrices are sparse, which can be utilized in software tools, such as MATLAB, and second, the matrix $(\mathbf{C}_{oo}^{-1} + \mathbf{S}_{\mathrm{SIR}}^{T}\mathbf{H}_{0}^{T}\mathbf{C}_{ee}^{-1}\mathbf{H}_{0}\mathbf{S}_{\mathrm{SIR}})$ in Eq. (22) is symmetric and positive definite, which enables the use of efficient iterative



(a) Point target at 50 mm. (b) Point target 190 mm.

FIG. 9. Profile plots (max amplitude of A-scans) comparison of the SAFT processed (dash-dotted) and the MMSE filter processed measured data (solid line) for a transducer size of 16 mm.

methods such as the preconditioned conjugate gradient method.[26] Also, the filter matrix $\hat{\mathbf{K}}$ can be computed off-line and then only a sparse matrix-vector multiplication has to be performed.

VI. CONCLUSIONS

A new method for ultrasonic synthetic aperture imaging using finite-sized transducers has been presented in this paper. The proposed algorithm takes the form of a model based MMSE filter, formulated in the time domain and performing deconvolution of the SIRs.

A compact, linear, discrete two-dimensional model of pulse-echo measurements was developed using matrix formalism. The matrix formulation that was subsequently used for the derivation of the MMSE method enabled solving the complex filtering problem with spatially and temporally varying coefficients. The filter matrix used for compensating diffraction effects of finite-sized transducers was derived using the criterion minimizing the mean square error for the linear discrete model of the imaging system. The resulting algorithm takes the form of a spatiotemporal filter that compensates for the SIRs associated with a finite-sized transducer at every point of the processed image (B-scan).

It was shown that the filters in SAFT algorithm, which can also be seen as a spatiotemporal filter, are a special case of the MMSE filters obtained for high noise situations and point-like transducers. Therefore, the resolution in images processed using the MMSE method will always be higher than that obtained using the SAFT algorithm (as long as the transducer is finite and the noise is not too large). This was clearly proven by the results of the experiments showing that the size effects of the transducer were significantly reduced by the MMSE method in contrast to the SAFT algorithm.

In summary, the proposed MMSE method is a generalization of SAFT that enables the use of ultrasonic synthetic aperture imaging in the case of finite-sized transducers.

ACKNOWLEDGMENT

This work was sponsored by the Swedish Nuclear Fuel and Waste Management Co. (SKB).

APPENDIX A: DISCRETE FORMULATION OF SAFT

To achieve focus at image point (x,z) the SAFT algorithm performs time shifts and a summation of the received signals $u_o(x_n,t)$ measured at transducer positions x_n for all n in the synthetic aperture. The time shifts are simply calculated using the Pythagorean theorem and the operation is commonly expressed in a continuous time form as[10,27–29]

$$\hat{o}(x,z)=\sum_n \alpha_n u_o\left(x,\frac{2}{c_\mathrm{p}}\sqrt{(x-x_n)^2+z^2}\right). \tag{A1}$$

The apodization weights α_n can be added to reduce the sidelobe levels (at a cost of broader main lobe). Omitting α_n, that is, $\alpha_n \equiv 1$, results in a sinc-shaped beampattern.[10]

Normally the received signals are digitized. Using matrix formalism, with the measured data collected in a matrix $\mathbf{Y}$, the discretized SAFT algorithm (A1) can be expressed as

$$(\hat{\mathbf{O}})_{m,\tilde{n}}=\sum_{\substack{n=\tilde{n}-K \\ \{n|\tilde{n}-K\geqslant 0,\tilde{n}+K\leqslant N-1\}}}^{\tilde{n}+K} \alpha_n(\mathbf{Y})_{j,n}. \tag{A2}$$

The constraints $\{n|\tilde{n}-K\geqslant 0,\tilde{n}+K\leqslant N-1\}$ limit the synthetic aperture at the edges which otherwise consists of $2K+1$ terms. Note that since the signal is discretized the time shifts must be rounded toward the nearest sampling instance; that is, the propagation delay index j, given by

$$j\doteq\frac{2}{c_\mathrm{p}T_\mathrm{s}}\sqrt{(x_{\tilde{n}}-x_n)^2+z_m^2}, \tag{A3}$$

is rounded to the nearest integer, which introduces some errors. However, these errors can be reduced by performing an interpolation that can be performed both in the time and the frequency domain, or by oversampling the signal.

APPENDIX B: DISCRETE CONVOLUTIONS IN MATRIX FORM

1. Time-invariant convolutions in matrix form

Assume that $o(m)$ is given for $m=0,1,2,...,M-1$, and $h(m)$ is given for $m=0,1,2,...,p-1$, and that both are zero

otherwise. Also assume that $y(m)$ is obtained by means of a discrete one-dimensional convolution of $o(m)$ and $h(m)$, that is,

$$y(m)=h(m)*x(m)=\sum_{k=0}^{p-1} h(k)x(m-k)$$

$$=\sum_{k=0}^{M-1} h(m-k)x(k). \tag{B1}$$

This operation can be expressed in matrix form

$$\mathbf{y}=\begin{bmatrix} h(0) & 0 & & & \cdots & & 0 \\ h(1) & h(0) & \ddots & & & & \vdots \\ \vdots & \vdots & \ddots & & & & \vdots \\ h(p-1) & h(p-2) & \cdots & h(0) & 0 & \cdots & 0 \\ 0 & h(p-1) & & \cdots & h(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ 0 & \cdots & 0 & h(p-1) & \cdots & & h(0) \end{bmatrix}\mathbf{x}$$

$$=\mathbf{H}\mathbf{x}. \tag{B2}$$

Now consider a two-dimensional data matrix $\mathbf{Y}$ whose columns can be expressed as a convolution of the form (B2); that is, the nth column $\mathbf{y}_n$ is $\mathbf{y}_n=\mathbf{H}\mathbf{x}_n$, where $\mathbf{x}_n$ is the nth column in a matrix $\mathbf{X}$ containing the unconvolved data. All the above-mentioned convolutions can now be expressed using the following compact matrix notation:

$$\mathbf{y}=\begin{bmatrix} \mathbf{H} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H} \end{bmatrix}\mathbf{x}=\mathbf{H}_0\mathbf{x}, \tag{B3}$$

where $\mathbf{y}$ is redefined as $\mathbf{y}=\mathrm{vec}(\mathbf{Y})$ and $\mathbf{x}$ as $\mathbf{x}=\mathrm{vec}(\mathbf{X})$.

2. SIR convolutions in matrix form

The SIR convolutions can be expressed in a similar form as in Eq. (30). Here, the object function vector $\mathbf{o}_{\tilde{n}}$ corresponds to $\mathbf{x}_n$ and $\mathbf{P}_{d(\tilde{n},n)}$ corresponds to $\mathbf{H}$. Both $\mathbf{H}$ and $\mathbf{P}_{d(\tilde{n},n)}$ are lower triangular but $\mathbf{P}_{d(\tilde{n},n)}$ does not have the Toeplitz form as $\mathbf{H}$ since the SIRs are different for different image points.

[1] M. Fatemi and A. C. Kak, "Ultrasonic B-scan imaging: Theory of image formation and a technique for restoration," Ultrason. Imaging **2**, 1–47 (1980).

[2] D. Iraca, L. Landini, and L. Verrazzani, "A piece-wise spatial invariant model of ultrasonic image formation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 457–463 (1988).

[3] W. Vollmann, "Resolution enhancement of ultrasonic B-scan images by deconvolution," IEEE Trans. Sonics Ultrason. **29**, 78–83 (1982).

[4] D. Iraca, L. Landini, and L. Verrazzani, "Power spectrum equalization for ultrasonic image restoration," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 216–222 (1989).

[5] J. A. Seydel, *Ultrasonic Synthetic-aperture Focusing Techniques in NDT*, Research Techniques for Nondestructive Testing (Academic, New York, 1982).

[6] K. Nagai, "A new synthetic-aperture focusing method for ultrasonic B-scan imaging by the Fourier transform," IEEE Trans. Sonics Ultrason. **32**, 531–536 (1985).

[7] K. Mayer, R. Marklein, K. J. Langenberg, and T. Kreutter, "Three-dimensional imaging system based on Fourier transform synthetic aperture focusing technique," Ultrasonics **28**, 241–255 (1990).

[8] S. I. Nikolov and J. A. Jensen, "3D synthetic aperture imaging using a virtual source element in the elevation plane," in Proceedings of the IEEE International Ultrasonics Symposium, Puerto Rico, 2000.

[9] M. Karman and M. O'Donnell, "Synthetic aperture imaging for small scale systems," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 429–442 (1995).

[10] C. H. Frazier and W. D. O'Brien, Jr., "Synthetic aperture techniques with a virtual source element," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 196–207 (1998).

[11] M.-H. Bae and M.-K. Jeong, "A study of synthetic-aperture imaging with virtual source elements in B-mode ultrasound imaging systems," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1510–1519 (2000).

[12] D. Vray, C. Haas, T. Rastello, M. Krueger, E. Busseau, K. Schroeder, G. Gimenez, and H. Ermert, "Synthetic aperture-based beam compression for intravascular ultrasound imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 189–201 (2001).

[13] C. Passman and H. Ermert, "A 100 MHz ultrasound imaging system for dermatologic and ophthalmologic diagnostics," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 545–552 (1996).

[14] E. S. Ebbini and C. A. Cain, "Multiple-focus ultrasound phased-array pattern synthesis: Optimal driving-signal distributions for hyperthermia," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 540–548 (1989).

[15] M. Tanter, J.-F. Aubry, J. Gerber, J.-L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. I. Basic principles," J. Acoust. Soc. Am. **110**, 37–47 (2001).

[16] J.-F. Aubry, M. Tanter, J. Gerber, J.-L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. II. Experiments. Application to focusing through absorbing and reverberating media," J. Acoust. Soc. Am. **110**, 48–58 (2001).

[17] G. S. Kino, *Acoustic Waves: Devices, Imaging and Analog Signal Processing*, Prentice-Hall Signal Processing Series Vol. 6 (Prentice–Hall, Englewood Cliffs, NJ, 1987).

[18] A. Lhémery, "Impulse-response method to predict echo-responses from targets of complex geometry. I. Theory," J. Acoust. Soc. Am. **90**, 2799–2807 (1991).

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Estimation Theory Vol. I (Prentice–Hall, Englewood Cliffs, NJ, 1993).

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Detection Theory Vol. II (Prentice–Hall, Englewood Cliffs, NJ, 1998).

[21] M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the inverse filter," J. Acoust. Soc. Am. **108**, 223–234 (2000).

[22] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd ed. (McGraw–Hill, New York, 1984).

[23] G. L. Turin, "An introduction to matched filters," IRE Trans. Inf. Theory **6**, 311–329 (1960).

[24] P. Wu and T. Stepinski, "Spatial impulse response method for predicting pulse-echo fields from linear array with cylindrically concave surface," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1283–1297 (1999).

[25] H. Lasota and R. Salamon, "Application of time-space impulse responses to calculations of acoustic field in imaging systems," Acoust. Imaging **10**, 493–512 (1980).

[26] J. Nocetal and S. J. Wright, *Numerical Optimization* (Springer, Berlin, 1999).

[27] M. Lorenz, "Ultrasonic imaging for the characterization of defects in steel components," Ph.D. thesis, Delft University of Technology, 1993.

[28] R. Hannemann, "Modeling and imaging of elastodynamic wave fields in homogeneous anisotropic media," Ph.D. thesis, Kassel University, 2001.

[29] S. Nikolov, "Synthetic aperture tissue and flow ultrasonic imaging," Ph.D. thesis, Technical University of Denmark, 1993.

Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix

Claire Prada[a] and Jean-Louis Thomas
Laboratoire Ondes et Acoustique, Université Denis Diderot, UMR CNRS 7587, ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 20 May 2002; revised 19 February 2003; accepted 19 February 2003)

The D.O.R.T. method (French acronym for Decomposition of the Time Reversal Operator) is an active remote sensing technique using arrays of antennas for the detection and localization of scatterers [Prada *et al.*, J. Acoust. Soc. Am. **99**, 2067–2076 (1996)]. The analogy between the time reversal operator and the covariance matrix used for classical sources separation in passive remote sensing [Bienvenu *et al.*, IEEE Trans. ASSP **31**, 1235–1247 (1983)] is established. Then, an experiment of subwavelength detection and localization of point-like scatterers with a linear array of transducers is presented. Using classical estimators in reception like Maximum-Likelihood and Multiple Signal Characterization (MUSIC), two point-like scatterers separated by $\lambda/3$ and placed at 100λ from the array of transducers are resolved. In these experiments, the role of multiple scattering and the existence of additional eigenvectors associated with dipolar and monopolar radiation of each scatterer is discussed. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1568759]

PACS numbers: 43.60.Pt, 43.28.We, 43.20.Fn [JCB]

I. INTRODUCTION

Remote sensing techniques are developed for many applications in the field of electromagnetic or acoustic waves. Methods with superresolution properties, like MUSIC, have been the subject of an intense field of research in the framework of passive array processing. However, numerous applications involve the detection and the localization of scatterers and consequently require array processing adapted to active array, like the D.O.R.T. method described below. In this work we establish the analytical link between D.O.R.T. and methods developed for the passive detection of sources that are based on an eigendecomposition of a covariance matrix. This formal demonstration is illustrated by an experiment in which subwavelength resolution is obtained.

Active transducer arrays with electronic focusing and steering capabilities are now commonly used in medical imaging; they have also become available for nondestructive evaluation and even in underwater acoustics. In the Laboratoire Ondes et Acoustique, a new adaptive detection technique, the D.O.R.T. method was implemented to optimize the use of such arrays. The D.O.R.T. method (a French acronym for ''Décomposition de l'Opérateur de Retournement Temporel) is an active detection technique that efficiently detects and separates the responses of several scatterers in homogeneous or heterogeneous media,[1–3] even in the presence of reverberation.[4,5] It was applied to the detection of flaws in nondestructive testing,[6,7] target detection and identification in underwater acoustics, and in electromagnetism as the first stage in the resolution of an inverse problem.[8]

This method is based on the singular value decomposition of the matrix $\mathbf{K}(\omega)$ of the interelement responses of the array at frequency ω. The processing assumes only linearity and time shift invariance of the propagation. This decomposition provides the eigenvector decomposition of the product ${}^t\mathbf{K}^*(\omega)\mathbf{K}(\omega)$, the so-called time reversal operator. Hence, this processing, like time reversal or phase conjugation, is adaptive to phase aberrations induced by heterogeneities in the medium. It was shown that when the scatterers are small compared to the wavelength, the number of significant singular values is, in general, equal to the number of scatterers. Furthermore, for well-resolved scatterers of distinct apparent reflectivities, the associated singular vector is the response of one scatterer to the array. In fact, Chambers[11] has shown analytically that for a point-like scatterer, there are four singular values (three for a 2-D configuration) of which three are very weak in most configurations. When the scatterers have sizes comparable to or larger than the wavelength, there is a whole set of singular values and singular vectors associated either with the specular part or with resonances of the scattered wave.[12]

In the field of source detection, sophisticated array signal processing was developed to estimate the number, direction, and signal intensity of sources. Among them, the signal subspace-type techniques were given great attention. Such methods start with an estimate of the data covariance matrix deduced from the impinging wavefront on the receiving array. For noncorrelated sources, the decomposition of this matrix into a source space and a noise space provides the number of sources. Then, their directions and strengths are obtained by various estimators applied to the source space. These processings are employed either in RADAR and SONAR for the detection and separation of sources.[13,14]

In a first stage, the connection between the D.O.R.T. method and classical array processing is established by identifying the scatterers with secondary sources and showing that the time reversal operator can be written as a covariance matrix like the one introduced in passive source detection. In

[a]Electronic mail: claire.prada-julia@espci.fr

a second stage, classical estimators in reception like Maximum-Likelihood (ML) and MUltiple SIgnal Characterization (MUSIC), will be applied in order to evaluate the resolving power of this active method. The possibility of applying MUSIC processing in reception was shown independently by Devaney[15] and also discussed by Cheney.[16]

A 2-D experiment with a linear array and two identical wires ten times thinner than the wavelength will illustrate the resolution and the precision of the method in the presence of noise. It will be shown that the method can separate and locate two scatterers positioned less than half a wavelength apart. Furthermore, this will give an experimental evidence that each target is associated with a 3-D signal space, as predicted by Chambers, and will also illustrate how multiple scattering affects the singular values distribution for close scatterers.

II. THE D.O.R.T. METHOD

To recall the D.O.R.T. method in a general manner, a propagating medium with point-like scatterers is considered. The medium may be heterogeneous with gradual variation of density and sound speed. An array of M transmitters and an array of L receivers are used to detect these scatterers. Assuming time shift invariance and linearity, any transmit–receive process performed from array #1 of M transmitters to array #2 of L receivers, is described by the $L\times M$ inter-element impulse responses $k_{lm}(t)$. The expression of $r_l(t)$, the signal received on element number l, is the following:

$$r_l(t)=\sum_{m=1}^{M} k_{lm}(t)\otimes e_m(t)+b_l(t), \tag{1}$$

where $e_m(t)$ is the signal applied to element number m and $b_l(t)$ is noise. A Fourier transform of Eq. (1) leads to

$$R(\omega)=\mathbf{K}(\omega)E(\omega)+B(\omega), \tag{2}$$

where $R(\omega)$ is the received vector signal, $\mathbf{K}(\omega)$ is the transfer matrix from transmit array to receive array, $E(\omega)$ is the emitted vector and $B(\omega)$ a noise vector.

The transfer matrix $\mathbf{K}(\omega)$ can be measured by transmitting successively a set of N independent vectors. The Singular Value Decomposition (SVD) of this complex matrix is

$$\mathbf{K}(\omega)=\mathbf{U}(\omega)\Sigma(\omega)^t\mathbf{V}^*(\omega), \tag{3}$$

where $\Sigma(\omega)$ is a real diagonal matrix of singular values, $\mathbf{U}(\omega)$ and $\mathbf{V}(\omega)$ are unitary matrices.

The physical meaning of this singular value decomposition is illuminated through the analysis of the iterative time reversal process. This process has been widely described[1–3] for a single array of transmit and receive transducers and generalized to separate arrays[4] and the time reversal operator was defined in a general manner as ${}^t\mathbf{K}^*(\omega)\mathbf{K}(\omega)$. Once the matrix $\mathbf{K}(\omega)$ is measured, the processing is purely numerical so that the role of the two arrays may be switched. At this stage, only the reciprocity principle, the time shift invariance, and the linearity of the medium are used. The SVD of $\mathbf{K}(\omega)$ can now be interpreted: the eigenvalues of ${}^t\mathbf{K}^*(\omega)\mathbf{K}(\omega)$ are the squares of the singular values of $\mathbf{K}(\omega)$ and its eigenvectors are the columns of $\mathbf{V}(\omega)$. In consequence, the singular vectors of $\mathbf{K}(\omega)$ are the invariants of the iterative time reversal process. These invariants are quite well known for ''Rayleigh''-type scatterers and single scattering. In this case, the transfer matrix $\mathbf{K}(\omega)$ is the product of three matrices, $\mathbf{H}_2(\omega)$, $\mathbf{D}(\omega)$, and $\mathbf{H}_1(\omega)$, respectively, modeling (1) the propagation from the transmit array toward the scatterers, (2) the scattering, and (3) the propagation from the scatterers toward the receive array:

$$\mathbf{K}(\omega)=\mathbf{H}_1(\omega)\mathbf{D}(\omega)\mathbf{H}_2(\omega). \tag{4}$$

From this equation, one can see that the rank of the transfer matrix is less than or equal to the number of scatterers. If multiple scattering is negligible then $\mathbf{D}(\omega)$ is a diagonal matrix. Furthermore, if the scatterers have distinct reflectivities and are ideally resolved by both transmit and receive arrays, then there is a one-to-one correspondence between the scatterers and the singular vectors. In other words, each vector corresponds to the response of one scatterer to the array.

In fact, the assumption of isotropic scattering is only true for a point-like discontinuity in compressibility. In general, a scatterer is also associated with a density variation and a recent study[11] has shown that four eigenvalues of the time reversal operator are associated with each scatterer. However, for hard materials, if the scatterer is much smaller than a wavelength, then three of the eigenvalues are much smaller than the dominant one. This is why the multiple eigenstates have not been observed in experiments to date.

Following Eq. (4), the time reversal operator is now written as

$${}^t\mathbf{K}^*(\omega)\mathbf{K}(\omega)={}^t\mathbf{H}_2^*(\omega)\mathbf{D}^*(\omega){}^t\mathbf{H}_1^*(\omega)\mathbf{H}_1(\omega)\mathbf{D}(\omega)\mathbf{H}_2(\omega). \tag{5}$$

We will see below how this operator can be considered as an estimate of a covariance matrix of the kind introduced in passive source detection.

III. THE COVARIANCE MATRIX AND THE TIME REVERSAL OPERATOR

Before discussing the analogy between the time reversal operator and the covariance matrix, the construction of the covariant matrix in passive source detection is recalled.

A. The covariance matrix for passive source detection technique

In usual descriptions of source detection techniques, a distribution of unknown sources and an array of receivers are considered.[9,10] The problem to be solved is to find the number of sources and their directions.

Let $(s_1(t),\ldots,s_d(t))$ be the signals emitted by d unknown sources, and let $(r_1(t),\ldots,r_L(t))$ be the signals measured on an array of L receivers. Assuming that the propagation is linear and time invariant, the signal measured on receiver number l can be written as

$$r_l(t)=\sum_i h_{li}(t)\otimes s_i(t)+b(t), \tag{6}$$

where, for $1\leqslant i\leqslant d$ and $1\leqslant l\leqslant L$, $h_{li}(t)$ is the response from source number i to receiver number l. In the frequency domain, Eq. (6) can be written in matrix form as

$$R(\omega)=\mathbf{H}(\omega)S(\omega)+B(\omega), \tag{7}$$

where $R(\omega)$ is the receive vector signal, $\mathbf{H}(\omega)$ is the transfer matrix from the source onto the array, $S(\omega)$ is the source vector, and $B(\omega)$ the noise vector. If M realizations of the receive vector are measured $[R_m(\omega),1\leqslant m\leqslant M]$, an estimate of the covariance matrix can be obtained by averaging of the corresponding matrices $\mathbf{C}_m(\omega)=R_m(\omega)^t R_m(\omega)^*$.

Assuming that the medium and source positions are time invariant and that the noise is not correlated to the sources, the covariance matrix is expressed from the source covariance matrix as

$$\langle \mathbf{C}(\omega)\rangle=\mathbf{H}(\omega)\langle S(\omega)\,{}^tS(\omega)^*\rangle\,{}^t\mathbf{H}(\omega)^* + \langle B(\omega)\,{}^tB(\omega)^*\rangle. \tag{8}$$

If the sources are uncorrelated, the matrix $\langle S(\omega)\,{}^tS(\omega)^*\rangle$ is diagonal. If the elements of the noise vector $B(\omega)$ are also uncorrelated, mean zero and variance σ, then the noise term is

$$\langle B(\omega)\,{}^tB(\omega)^*\rangle=\sigma^2\mathbf{I}. \tag{9}$$

The next stage of spectral methods is to compute an eigendecomposition of the covariance matrix $\langle \mathbf{C}(\omega)\rangle$. If the responses from the sources are orthogonal vectors, then the preceding expression provides the diagonalization of the covariance matrix. The eigenvalues are the following:

$$\lambda_i=\langle |S_i(\omega)|^2\rangle\sum_{l=1}^{L}|H_{il}(\omega)|^2+\sigma^2. \tag{10}$$

The eigenvalues are proportional to the average intensity of the sources and the eigenvectors are the vectors H_i, i.e., the frequency response from the source to the array. However, in a general case, the responses from the sources to the array are not orthogonal vectors.

B. Link between the time reversal operator and the covariance matrix

We now reconsider the measurement of the transfer matrix in a scattering experiment with ''Rayleigh-type'' scatterers. We use the same notations as in Eqs. (4) and (5).

Insonifying the medium with a pulse on transducer n, the received vector signal is the nth column of $\mathbf{K}$ and is expressed as

$$R_n(\omega)=\mathbf{H}_1(\omega)\mathbf{D}(\omega)H_{2n}(\omega)+B_n(\omega), \tag{11}$$

where $H_{2n}(\omega)$ is the nth column of $\mathbf{H}_2(\omega)$ and $B_n(\omega)$ is the noise vector.

The vector $\mathbf{D}(\omega)H_{2n}(\omega)$ can be considered as one realization of the source vector $S(\omega)$ of Eq. (7). The measurement of the N columns of $\mathbf{K}$ is analogous to N measurements of the source signal in passive detection (Sec. III A). If the aperture of the array and the number of elements are large enough, there will be enough diversity in the measurements to decorrelate these secondary sources.

Though the responses of the scatterers $R_n(\omega)$ are purely deterministic, they can be used to built a ''pseudocovariance matrix'' as follows:

$$\tilde{\mathbf{C}}(\omega)=\sum_n R_n(\omega)^t R_n^*(\omega). \tag{12}$$

Using the expression of $R_n(\omega)$ in Eq. (11), $\tilde{C}(\omega)$ is written as

$$\begin{aligned}\tilde{\mathbf{C}}(\omega)=&\mathbf{H}_1(\omega)\mathbf{D}(\omega)\\ &\times\left(\sum_n H_{2n}(\omega)^t H_{2n}^*(\omega)\right)\mathbf{D}^*(\omega)^t\mathbf{H}_1^*(\omega)\\ &+\mathbf{H}_1(\omega)\mathbf{D}(\omega)\left(\sum_n H_{2n}(\omega)^t B_n^*(\omega)\right)\\ &+\left(\sum_n B_n(\omega)^t H_{2n}^*(\omega)\right)\mathbf{D}^*(\omega)^t\mathbf{H}_1^*(\omega)\\ &+\left(\sum_n B_n(\omega)^t B_n^*(\omega)\right).\end{aligned} \tag{13}$$

If the noise vector is not correlated to the scatterer responses, and if the number of measurements is large enough, the cross-terms vanish. Using Eq. (9) the expression simplifies to

$$\tilde{\mathbf{C}}(\omega)\cong\mathbf{H}_1(\omega)\mathbf{D}(\omega)^t\mathbf{H}_2(\omega)\mathbf{H}_2^*(\omega)\mathbf{D}^*(\omega)^t\mathbf{H}_1^*(\omega) + n\sigma^2\mathbf{I}. \tag{14}$$

In the absence of noise ($\sigma=0$), this last expression is exactly the time reversal operator $\mathbf{K}(\omega)^t\mathbf{K}^*(\omega)$ [see Eq. (5)]. It can also be compared with Eq. (8), in which the source covariance matrix $\langle S(\omega)^tS(\omega)^*\rangle$ is replaced by the matrix $\Sigma_n\mathbf{D}(\omega)^tH_{2n}(\omega)H_{2n}^*(\omega)\mathbf{D}^*(\omega)=\mathbf{D}(\omega)^t\mathbf{H}_2(\omega)\mathbf{H}_2^*(\omega)\mathbf{D}^*(\omega)$, where the average is taken over the different emissions. In general, when the number of scatterers is less than the number of emitters, the rank of this last matrix is equal to the number of scatterers.

The next stage consists in computing the eigenvalues and eigenvectors of the covariance matrix, which is equivalent to computing an SVD of the transfer matrix $\mathbf{K}$. This decomposition is a way of sorting the data from most dominant to the weakest contribution.

IV. DETECTION PROCESSING AND SUPER-RESOLUTION

In order to improve the resolution of the D.O.R.T. method, we have tried two estimators developed in classical array processing using the covariance matrix: Maximum Likelihood (ML) and MUSIC. To perform this analysis the vector signal $G(M)$ received by the array from a scatterer located at point M has to be modeled. In our model we choose the Green's function for free space.

The Maximum Likelihood estimator looks for the maxima of the function $I_{\mathrm{ML}}(M)$:

$$I_{\mathrm{ML}}(M)=\frac{1}{{}^tG^*(M)\mathbf{C}^{-1}G(M)}. \tag{15}$$

As the inverse of the covariance matrix may be computed by using the SVD of $\mathbf{C}$, Eq. (15) becomes

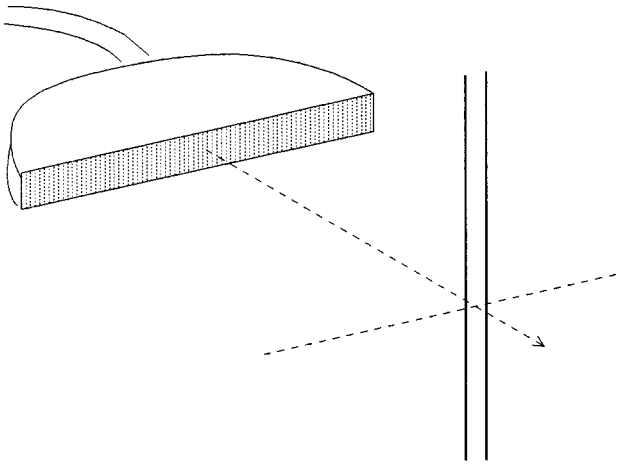


FIG. 1. Experimental setup.

$$I_{\mathrm{ML}}(M)=\frac{1}{{}^{t}G^{*}(M)\mathbf{U}\mathbf{\Lambda}^{-1}\,{}^{t}\mathbf{U}^{*}G(M)}$$

$$=\frac{1}{\sum_{i=1}^{L}\frac{1}{\lambda_i}|\langle\mathbf{U}_i|G(M)\rangle|^2}, \tag{16}$$

where $\Lambda=\Sigma\Sigma$ is the diagonal matrix composed of the eigenvalues of $C(\omega)$. A great advantage of this method is that it does not require the knowledge of the source space dimension.

If the dimension of the signal space, p, is known or can be estimated from the singular value distribution, then the MUSIC algorithm applies. The estimator is then

$$I_{\mathrm{MU}}(M)=\frac{1}{1-\Sigma_{i=1}^{p}|\langle\mathbf{U}_i|G(M)\rangle|^2}$$

$$=\frac{1}{\sum_{i=p+1}^{L}|\langle\mathbf{U}_i|G(M)\rangle|^2}. \tag{17}$$

This function is the inverse of the square of the Euclidean distance from the Green's vector $G(M)$ to the signal space. These two estimators are very similar, but MUSIC gives more importance to the eigenvector of the noise space.

It appears that these two estimators are nonlinear combinations of the back-propagation of each singular vector, $I_v(M)=|\langle\mathbf{V}|G(M)\rangle|$, used in our preceding papers.[3]

In all of these three methods, the choice of the Green's function is crucial and any *a priori* knowledge introduced to improve the model is worth trying.

V. EXPERIMENTAL RESULTS

The experimental results are obtained in a 2-D geometry with a linear transmit and receive array and two point-like scatterers (Fig. 1). A single plane linear array of 128 piezoelectric transducers is used for the transmission and the reception. The array pitch is 0.5 mm and the central frequency is 1.5 MHz with 60% bandwidth. This half-wavelength pitch allows a very good spatial sampling of the ultrasonic field.

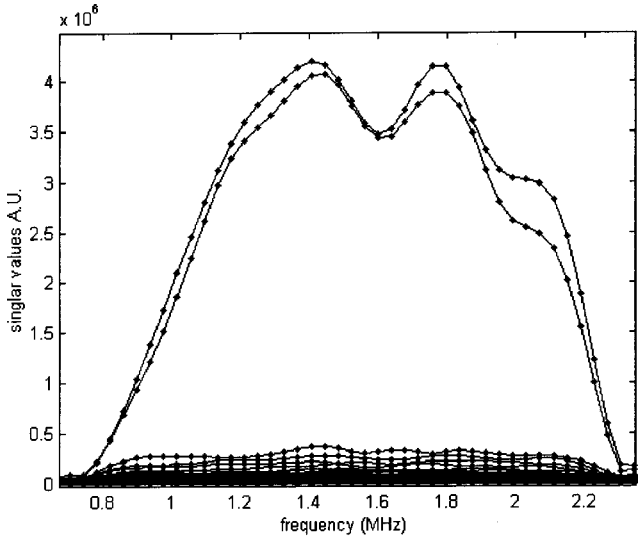


FIG. 2. Singular values versus frequency for a wire separation distance $d=3.5$ mm (arbitrary units).

The height of the piezoelectric transducer is 10 mm. The received signals are sampled at a frequency of 20 MHz. The scatterers are two copper wires of 0.1 mm diameter, which is approximately a tenth of a wavelength. The wires are placed perpendicular to the array at a range of 60 mm. They can be considered as point-like in the plane of insonification. This setting and the small size of the scatterers ensures a Rayleigh-like scattering for the whole bandwidth of the transducers. A translation system was built to vary precisely the distance between the wires. The two wires are stretched on a rigid frame. The fasteners of one of them are mobile on slides and the distance between the wires is then regulated with block gauges.

A. Acquisition of the array response matrix

The most natural way to acquire the array response matrix is to use the canonical basis, that is to say, to acquire separately the response of each pair of elements in the array. However, in our experiment the scattered energy is weak as each scatterer is small compared to the wavelength. In order to improve the signal to noise ratio, the array was driven using the more energetic Walsh function basis,[17] an orthonormal basis in which all the elements are excited in each transmission. However, the measured matrix is transformed to the canonical basis before computing its SVD. This procedure allows a significant increase of signal to noise ratio (around 20 dB for an array of 128 elements). Acquisition of the array response matrix was done for several wire separation distances **d** ranging from 0.3 to 3.5 mm.

B. Result of the decomposition for well-resolved scatterers $d=3.5$ mm

In the case of two targets separated by a distance much greater than the array point spread function, the corresponding Green's functions are almost orthogonal. The SVD of the matrix **K** provides a decomposition of the received signal space into a signal subspace of rank 2 and a noise subspace of rank 126. Figure 2 displays the amplitude of the singular values for all frequencies in the bandwidth of the transducers. As the two wires are identical and well resolved, the two

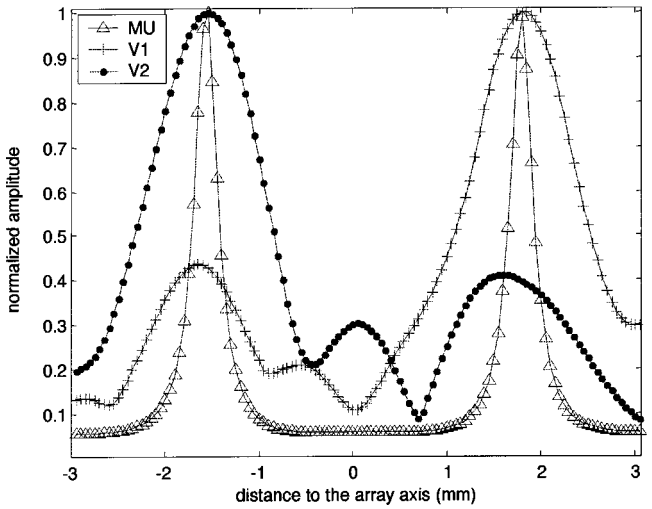


FIG. 3. Wires spaced 3.5 mm, image obtained by back-propagation of the two main eigenvectors (V1 and V2), compared to the image obtained by MUSIC at frequency 1.52 MHz.

highest singular values have almost the same amplitude. To locate the two scatterers, one may compute the scalar product of the singular vectors V_1 and V_2 with the Green's function $G(M)$ for all points M in the area of interest. This treatment is a focusing in receive mode also called back-propagation or virtual time reversal. This treatment in receive mode provides the position of each scatterer (Fig. 3, ● and +). In this case, each singular vector focuses onto both scatterers. In fact, even if the two scatterers are well resolved by the array, the symmetry of the experiment causes a significant coupling between the scatterers' responses that has a strong influence on the singular vectors. In several papers, we have used this back-propagation to make different images with the dominant singular vectors.[1–5]

Using the MUSIC estimator, a higher resolution is achieved. The main lobes of the image provided by MUSIC is five times thinner than with conventional focusing. However, the evaluation of the resolution limit of these methods can only be achieved by studying less distant scatterers. This will be pursued in Sec. V D.

C. The singular values as a function of the distance between the scatterers

When the distance between the two wires decreases to zero, the correlation between the two corresponding Green's functions increases. This coupling modifies the singular values. In the present case of two identical scatterers of reflectivity C in a symmetrical geometry, the first two singular values can be expressed:

$$s_{\pm}=C\left(\sum_{l=1}^{N}|H_{1l}|^2\pm\sum_{l=1}^{N}H_{1l}H_{2l}^{*}\right). \tag{18}$$

We see that if the two scatterers are very close, $H_1\approx H_2=H$ and $H_1-H_2=dH$, and Eq. (18) can be approximated by

$$s_{+}=2C\|H\|^2 \quad \text{and} \quad s_{-}=2C\langle H|dH\rangle.$$

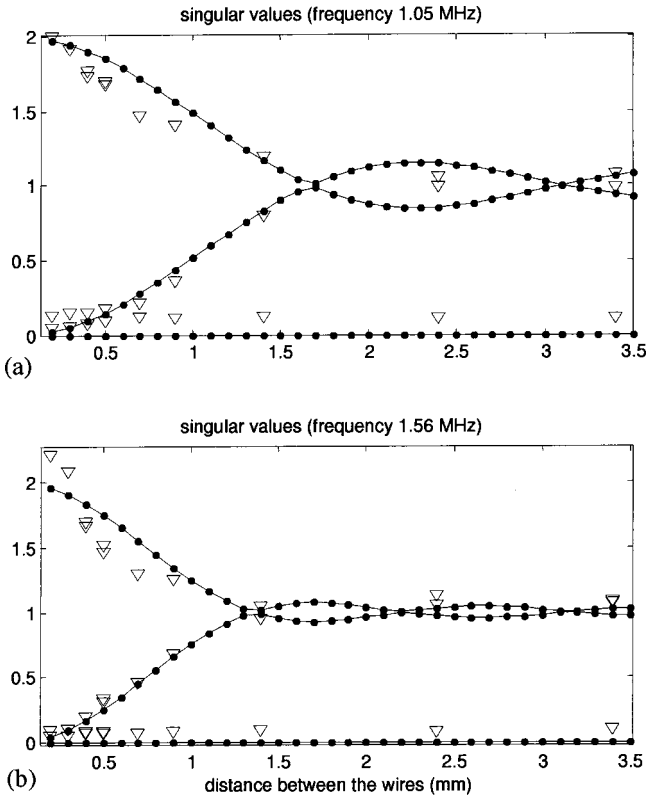


FIG. 4. Experimental (triangle) three main singular values versus the wire separation distance compared with a model (solid line with full circle); (a) frequency 1.05 MHz and (b) frequency 1.56 MHz.

When the scatterers get closer, s_+ converges to twice the apparent reflectivity of the scatterers, while s_- converges toward 0.

A simple model using $\exp(jkr)/\sqrt{r}$ as a surrogate for the 2-D free space Green's function, and the assumption that the transducers are point-like was used to calculate the theoretical singular values of the array response matrix. These theoretical values are compared to the experimental ones in Fig. 4.

The third experimental singular value shows the noise level. It appears to be independent of the distances between the wires. We see that when the distance between the wires reaches 0.4 mm (or 0.4 wavelengths) the second singular value reaches the noise singular values. Looking only at the singular values, one may conclude that wires closer than half a wavelength cannot be distinguished. However, as will be shown later, higher resolution can be achieved through appropriate processing in reception.

Furthermore, for distances between the wires smaller than the wavelength, there is some discrepancy between the theoretical and experimental singular values. Introducing multiple scattering in the simulation, a similar effect is observed on the first singular value (Fig. 5). It is interesting to note that multiple scattering does not change the number of significant singular values. Though multiple scattering induces a global phase shift and global amplitude factor on the wave front emerging from the two wires, the wave impinging on array elements is still a linear combination of these two

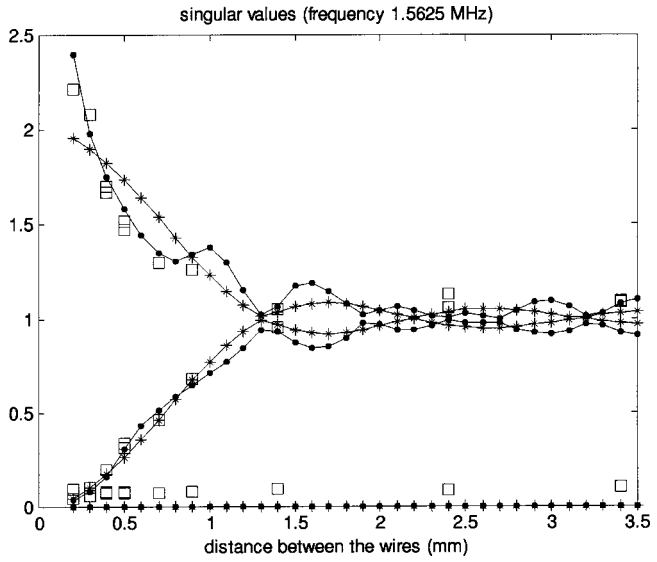


FIG. 5. Three main singular values versus the wire separation distance compared with a model, taking into account multiple scattering at frequency 1.56 MHz. Experiment: square symbol, simple model: solid line with stars, model taking into account multiple scattering: solid line with dots.

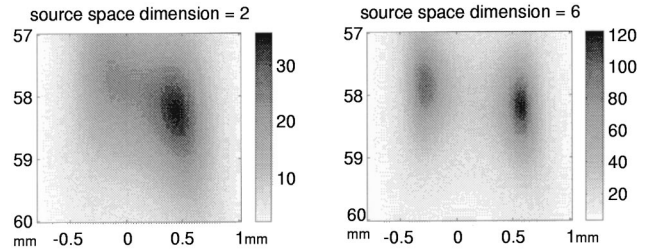


FIG. 6. Image obtained with the MUSIC criterion assuming a signal space of dimension 2 (left) and of dimension 6 (right).

wave fronts, consequently, the rank of the matrix **K** is unchanged.

D. Detection and localization for subwavelength separation

The following experimental results demonstrate the super resolution property of the Maximum Likelihood and MUSIC methods in the case of active array and scatterers. The analysis is performed for frequencies between 0.8 and

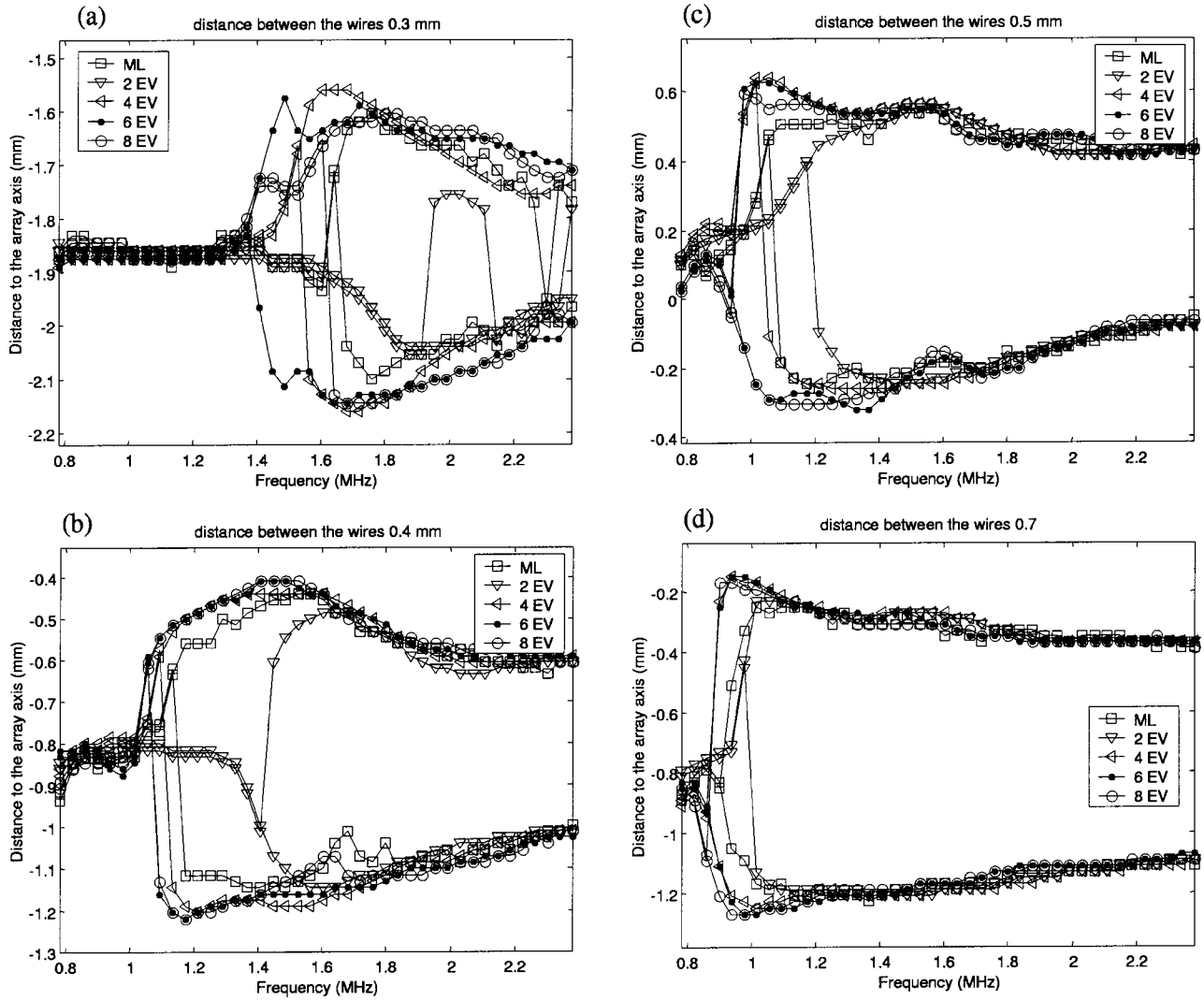


FIG. 7. Estimated lateral coordinates for the Maximum Likelihood (ML) and for the MUSIC algorithm assuming the signal space has dimension 2, 4, 6, and 8 (2, 4, 6, and 8 eV). From top to bottom the distances between the wires are 0.3, 0.4, 0.5, and 0.7 mm.

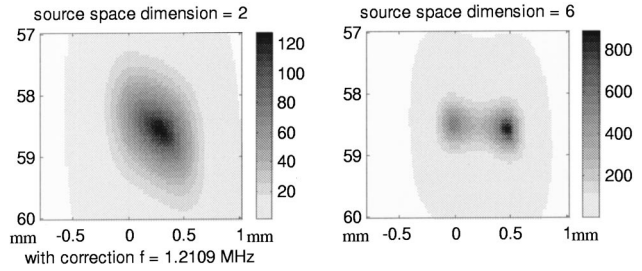


FIG. 8. Image obtained with MUSIC criterion for a signal space of dimension 2 (left) and of dimension 6 (right).

2.4 MHz (λ ranging from 0.6 to 1.9 mm) and for wire axis separation distances $d=0.3$, 0.4, 0.5, and 0.7 mm.

The first example examines the case of $d=0.5$ mm. The MUSIC estimator is computed in a small area around the location of the wires at frequency 1.34 MHz and assuming a signal subspace of dimension 2, Fig. 6 left. The image presents two maxima located around 58 mm in range and separated by a distance of around 0.6 mm. The criterion is then computed using the six dominant singular vectors, Fig. 6 right. A significant resolution gain can be observed. This result suggests that the signal subspace has a dimension higher than 2, as predicted by Chambers;[11] more evidence of this is presented later.

To get better estimates of the resolution limit of the MUSIC and ML methods, this processing was repeated for all frequencies and for the four wire separation distances. Furthermore, the MUSIC criterion was computed for a signal subspace of dimension 2, 4, 6, or 8. Figure 7 presents the lateral positions of the maxima for each estimator. This last figure illustrates the super-resolution property of these methods. The best results are obtained with MUSIC for a signal subspace of dimension 6. The resolution frequency threshold is 1.5, 1.1, 1, and 0.9 MHz for expected distances between the wires of 0.3, 0.4, 0.5, and 0.7 mm, respectively. As expected, MUSIC and ML provide much better resolution than classical focusing, for which the point spread function of the array is 1 mm at 1.5 MHz. Furthermore, the resolution limit computed as the ratio between the wavelength and the wire separation distance is, respectively, 3.3, 3.4, 3 and 2.4. For $d=0.5$ mm and $d=0.7$ mm, this resolution threshold is limited primarily by the bandwidth of the transducer. Indeed, the sensitivity below 1 MHz is very weak so that the difference between the second singular value and the noise singular values is small (Fig. 2). For $d=0.3$ mm and $d=0.4$ mm, the separation occurs when the two wires are separated by less than $\lambda/3$. This result is remarkable, as the array is located about 60 wavelengths from the scatterers, where evanescent waves are negligible.

The sharpening of the criterion shown in Fig. 6 and the fact that the best resolution is reached with a signal subspace of dimension 6 (Fig. 7), indicate that each scatterer is associated with a 3-D signal space, as predicted by Chambers[11] in a 2-D configuration. These three dimensions come from the fact that, in general, the incident field is scattered by heterogenities in both density and compressibility. Heterogeneity in compressibility gives rise to monopolar radiation that leads to a lone isotropic singular vector. In contrast, heterogeneity

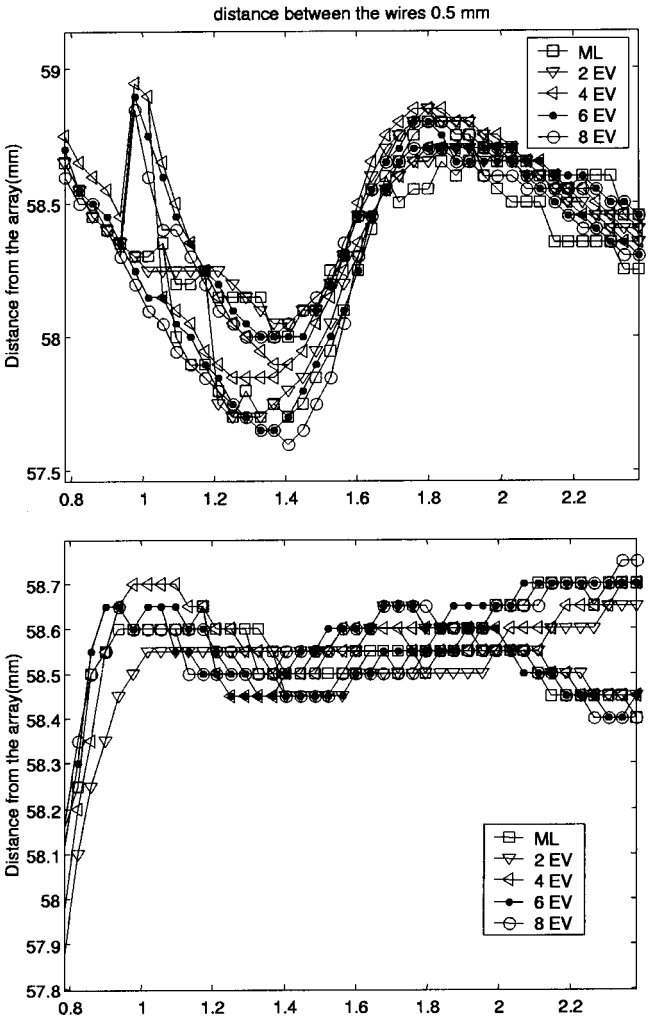


FIG. 9. Distance from the wires to the array (top: no correction; bottom: corrected matrix).

of density gives rise to dipolar radiation. This nonisotropic behavior is responsible for additional singular vectors. This is the first experimental evidence of the existence of these additional singular values.

E. Improvement of the accuracy by correction of the transfer matrix

One important drawback of these two super-resolution methods is the loss of accuracy in the estimated location of the scatterers when the wavelength increases, i.e., the correlation between the responses of the two scatterers increases. This deterioration of the performance is partly accounted for by increased sensitivity to the discrepancies between the true response of each scatterer and the model used. Previously, the array elements were assumed point-like and identical. The model has two main imperfections: first, the sensitivity varies from one element to the other; second, the finite size of the array elements leads to averaging of the incident field and induces amplitude and phase modulation that vary with both the frequency and direction of scattering In the following an empirical correction is applied on the experimental measurements to take into account these array imperfections. As a correction we use the first eigenvector V of the matrix

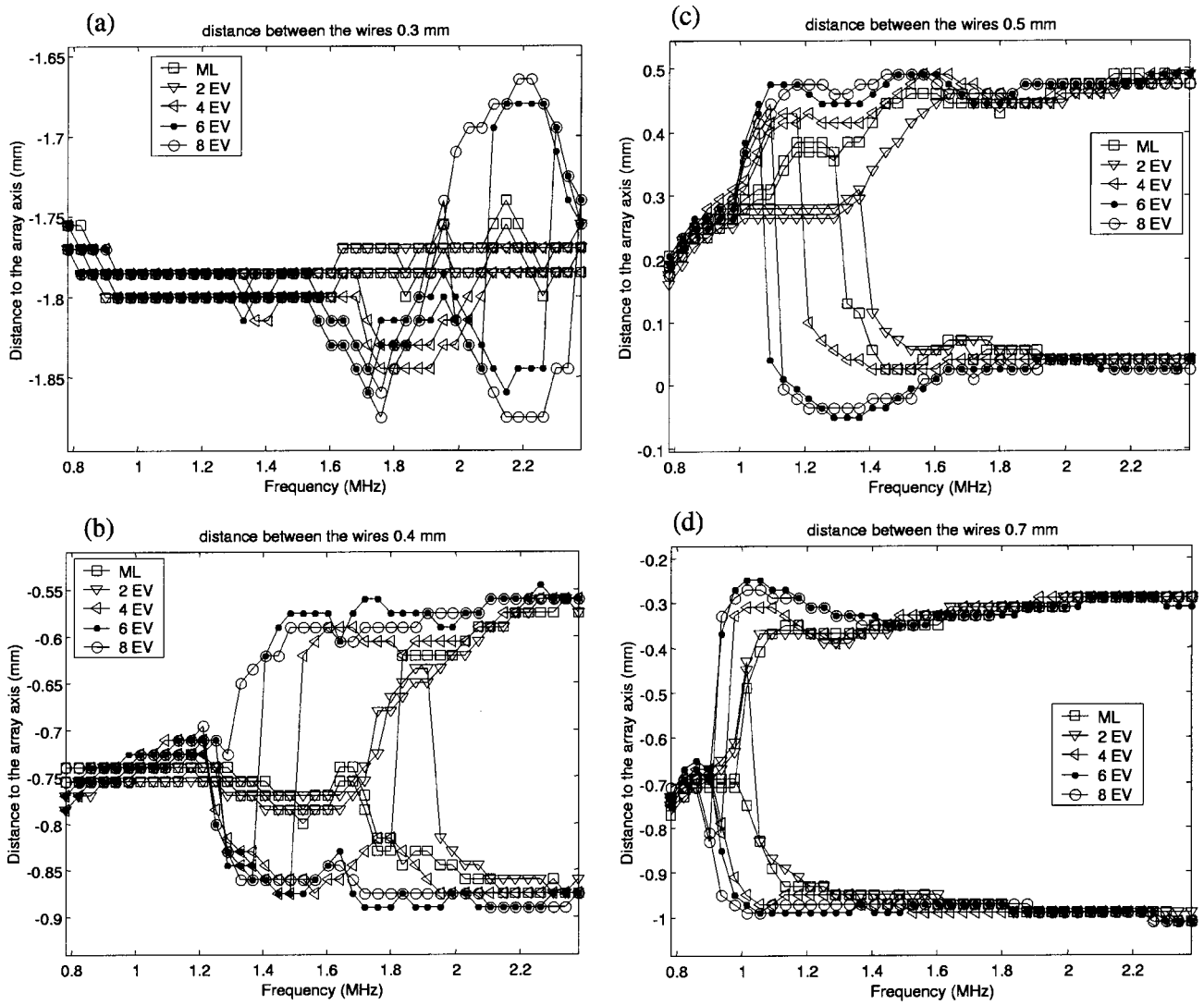


FIG. 10. Estimated coordinate for the corrected matrix using maximum likelihood (ML) and the MUSIC algorithm, assuming the signal space is dimension 2, 4, 6, and 8 (2, 4, 6, and 8 eV). From top to bottom the distances between the wires are 0.3, 0.4, 0.5, and 0.7 mm.

measured for a single wire. The error between this eigenvector and the estimated response Ve is used to compensate each column of the array response matrix $K(\omega)$. More precisely, we replace $K_{ij}(\omega)$ by $K_{ij}(\omega)/V_i(\omega)^* Ve_i(\omega)$, for all pairs i,j. The SVD of the corrected matrix is performed and, as was done to the original matrix, the MUSIC estimator is calculated for signal spaces of dimensions 2 and 6 (Fig. 8). The level of the peak is 7 times higher than before correction (Fig. 6) and the resolution of the image is much better. The improvement is also significant for range localization, as can be seen in Fig. 9. The estimated range is 58.6 mm with a fluctuation of ± 0.15 mm within the whole bandwidth. Without correction, the amplitude fluctuation was larger by at least a factor of 3. Similarly, fluctuations in lateral position are decreased significantly, and now vary by ± 0.025 mm around the mean position (Fig. 10). This is a quarter of the wire diameter.

Our correction, however, decreases the resolution power. The two wires 0.3 mm away are no longer resolved and the resolution threshold is, respectively, 1.4, 1.1, and 0.9 MHz for expected distances between the wires of 0.4, 0.5, and 0.7 mm, respectively. The ratio of wavelength to wire separation is, respectively, 2.7, 2.7, and 2.4. Again the bandwidth is the limiting factor at 0.9 MHz. The experimental Green's function, used to correct our model is necessarily noisy. This produces noise in our empirical model and we believe is responsible for this loss of resolution.

VI. CONCLUSION

In this paper we establish the analogy between the time reversal operator and the pseudocovariance matrix used in high resolution subspace techniques for source detection and localization with passive sensor arrays. As a consequence, the D.O.R.T. method appears as a generalization to active arrays of classical source separation processing. The D.O.R.T. processing starts with the successive transmission of N independent signals resulting in N backscattered wave fronts from which the ''pseudocovariance matrix'' is constructed. The decorrelation of the scatterers' responses is obtained thanks to the spatial diversity of emissions achieved by the array.

An experiment conducted with a linear array of 128 antennas and 2 thin wires illustrates the resolution power of the technique in the farfield. Two estimators were used to determine the position of the wires: the MUSIC and the Maximum Likelihood estimators. Both methods provide resolution much finer than the point spread function, and the best resolution was achieved with the MUSIC algorithm assuming a signal space of dimension 6. The found resolution limit is $\lambda/3$, which is three times smaller than the point spread function (here λ) and beyond the classical diffraction limit $\lambda/2$. For such small distances multiple scattering becomes significant, and we observed that although this phenomenon modifies the eigenvalues distribution, it does not affect the localization process.

The D.O.R.T. method as the time reversal process is robust in heterogeneous media. However, when the acoustic properties of the medium are not constant, any estimator in reception like MUSIC or ML requires an accurate model of the medium. This new method has many potential applications in ultrasonic imaging such as, for example, the detection and localization of flaws in nondestructive evaluation and the detection of mines in underwater acoustics.

[1] C. Prada and M. Fink, ''Eigenmodes of the time reversal operator: a solution to selective focusing in multiple-target media,'' Wave Motion **20**, 151–163 (1994).

[2] C. Prada, J.-L. Thomas, and M. Fink, ''The iterative time reversal process: analysis of the convergence,'' J. Acoust. Soc. Am. **97**, 62–71 (1995).

[3] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, ''Decomposition of the time reversal operator: Application to detection and selective focusing on two scatterers,'' J. Acoust. Soc. Am. **99**, 2067–2076 (1996).

[4] C. Prada, M. Tanter, and M. Fink, ''Flaw detection in solid with the D.O.R.T. method,'' IEEE Ultrason. Symp. Proc. **1**, 681–686 (1997).

[5] E. Kerbrat, C. Prada, D. Cassereau, R. K. Ing, and M. Fink, ''Detection and imaging in complex media with the D.O.R.T. method,'' IEEE Ultrason. Symp. Proc. **1**, 779–783 (2000).

[6] N. Mordant, C. Prada, and M. Fink, ''Highly resolved detection and selective focusing in a waveguide using the D.O.R.T. method,'' J. Acoust. Soc. Am. **105**, 2634–2642 (1999).

[7] T. Yokoyama, T. Kikuchi, T. Tsuchiya, and A. Hasegawa, ''Detection and selective focusing on scatterers using decomposition of time reversal operator method in pekeris waveguide model,'' Jpn. J. Appl. Phys. **40**, 3822–3828 (2001).

[8] H. Tortel, G. Micolau, and M. Saillard, ''Decomposition of the time reversal operator for electromagnetic scattering,'' J. Electromagn. Waves Appl. **13**, 687–719 (1999).

[9] G. Bienvenu and L. Kopp, ''Optimality of high resolution array processing using the eigensystem approach,'' IEEE Trans. Acoust., Speech, Signal Process. **31**, 1235–1247 (1983).

[10] R. O. Schmidt, ''Multiple emitter location and signal parameter estimation,'' IEEE Trans. Antennas Propag. **34**, 276–280 (1989).

[11] D. H. Chambers and A. K. Gautesen, ''Time reversal for a single spherical scatterer,'' J. Acoust. Soc. Am. **109**, 2616–2624 (2001).

[12] S. Komilikis, C. Prada, and M. Fink, ''Characterization of extended objects with the D.O.R.T. method,'' IEEE Ultrason. Symp. Proc. **2**, 1401–1404 (1996).

[13] M. I. Skolnik, RADAR Handbook (McGraw-Hill, New York, 1990).

[14] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, ''An overview of matched Field Methods in Ocean Acoustics,'' IEEE J. Ocean. Eng. **18**, 401–424 (1993).

[15] A. J. Devaney, ''Computational time reversal and object localization'' J. Acoust. Soc. Am **110**, 2617 (2001); or A. J. Devaney, ''Super-resolution processing of multi-static data using time reversal and music,'' submitted to J. Acoust. Soc. Am.

[16] M. Cheney, ''The linear Sampling Method and the MUSIC Algorithm,'' Inverse Probl. **17**, 591–595 (2001).

[17] S. Azzaretti, D. Dotti, R. Lombardi, and D. Rossi, ''Echo-graphic images enhanced through Walsh functions,'' IEEE Ultrason. Symp. Proc. **1**, 675–678 (1997).

Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves

Christopher A. Shera[a)]
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 11 December 2002; revised 20 March 2003; accepted 24 March 2003)

Mammalian spontaneous otoacoustic emissions (SOAEs) have been suggested to arise by three different mechanisms. The local-oscillator model, dating back to the work of Thomas Gold, supposes that SOAEs arise through the local, autonomous oscillation of some cellular constituent of the organ of Corti (e.g., the ''active process'' underlying the cochlear amplifier). Two other models, by contrast, both suppose that SOAEs are a global collective phenomenon—cochlear standing waves created by multiple internal reflection—but differ on the nature of the proposed power source: Whereas the ''passive'' standing-wave model supposes that SOAEs are biological noise, passively amplified by cochlear standing-wave resonances acting as narrow-band nonlinear filters, the ''active'' standing-wave model supposes that standing-wave amplitudes are actively maintained by coherent wave amplification within the cochlea. Quantitative tests of key predictions that distinguish the local-oscillator and global standing-wave models are presented and shown to support the global standing-wave model. In addition to predicting the existence of multiple emissions with a characteristic minimum frequency spacing, the global standing-wave model accurately predicts the mean value of this spacing, its standard deviation, and its power-law dependence on SOAE frequency. Furthermore, the global standing-wave model accounts for the magnitude, sign, and frequency dependence of changes in SOAE frequency that result from modulations in middle-ear stiffness. Although some of these SOAE characteristics may be replicable through artful *ad hoc* adjustment of local-oscillator models, they all arise quite naturally in the standing-wave framework. Finally, the statistics of SOAE time waveforms demonstrate that SOAEs are coherent, amplitude-stabilized signals, as predicted by the active standing-wave model. Taken together, the results imply that SOAEs are amplitude-stabilized standing waves produced by the cochlea acting as a biological, hydromechanical analog of a laser oscillator. Contrary to recent claims, spontaneous emission of sound from the ear does not require the autonomous mechanical oscillation of its cellular constituents. [DOI: 10.1121/1.1575750]

PACS numbers: 43.64.Bt, 43.64.Kc, 43.64.Jb, 43.64.Ha, 43.25.Gf [BLM]

I. INTRODUCTION

Spontaneous otoacoustic emissions (SOAEs), first reported in 1979 (Kemp, 1979a; Wilson, 1980; Zurek, 1981), are commonly thought to arise through a mechanism whose essentials were described some 30 years earlier (Gold, 1948). Discussing the implications of his ''regeneration hypothesis'' that electromechanical feedback somehow counteracts the viscous damping in the cochlea, Gold noted that ''if the feedback ever exceeded the losses, then a resonant element [in the organ of Corti] would become self-oscillatory, and oscillations would build up [to] a level where linearity was not preserved.'' Despite the ''self-regulating mechanism,'' whose existence he postulated as necessary to control the amount of feedback, Gold suggested that

> ''We might expect that occasional disturbances would bring an element into the region of self-oscillation, when it is normally so close to this condition. If this occurred, then we should hear a clear note which would persist until the adjusting mechanism has regained control, or until the nervous sensitivity has decreased sufficiently.''

Fifty years after Gold's prediction, the vocabulary is new but the basic idea remains unchanged:

> ''Such a self-tuning mechanism provides a natural explanation for spontaneous emissions of sound from the ear. Normally, the low-amplitude vibration of the self-tuned critical oscillators would produce a faint hum. But if one of the motile systems were to have a faulty control mechanism, it might oscillate wildly, generating a shrill whistle.'' (Duke, 2002).

The descriptions quoted here share with many others in the literature a key feature: They identify the ''oscillating element'' responsible for spontaneous emission of sound as *local* to a particular place along the cochlea. Often the putative oscillating element is localized even further within a cochlear cross section and identified with ''certain cells'' or parts of cells (e.g., hair-cell stereocilia) within the organ of Corti. Martin and Hudspeth (2001), for example, adopt this view when arguing that ''unprovoked movements of some

a)Electronic mail: shera@epl.meei.harvard.edu

constituent of the ear's amplifier are expected to underlie the production of spontaneous otoacoustic emissions.''

A. Local oscillator or global standing-wave resonance?

Despite wide acceptance of the local-oscillator model of SOAEs, there exists another possibility for SOAE generation, a possibility first suggested by Kemp (1979a, 1979b) and subsequently elaborated in models of evoked otoacoustic emissions (e.g., Zwicker and Peisl, 1990; Zweig, 1991; Talmadge and Tubis, 1993; Zweig and Shera, 1995; Allen *et al.*, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 1999). These models of evoked emissions predict that mammalian SOAEs arise not via autonomous cellular oscillations but as cochlear standing-wave resonances. In this view, SOAEs result from multiple internal reflection of traveling-wave energy initiated either by sounds from the environment or by physiological noise.

The theory of reflection-source emissions (Shera and Zweig, 1993b; Zweig and Shera, 1995; Talmadge *et al.*, 1998), for example, predicts that backward-traveling cochlear waves are generated by the coherent scattering of forward-traveling waves off densely and randomly distributed perturbations in the mechanics of the cochlea. Because wavelets scattered near the peak of a forward-traveling wave have much larger amplitudes than those reflected elsewhere, the net reflected wave is dominated by scattering that occurs in the region about the response maximum. The resulting backward-traveling waves are then reflected by the impedance mismatch at the cochlear boundary with the middle ear, generating additional forward-traveling waves that subsequently undergo another round of coherent reflection near their characteristic places. At frequencies for which the total phase change due to round-trip wave travel is an integral number of cycles, standing waves can build up within the cochlea, which is then acting, in effect, as a tuned resonant cavity. Cochlear standing waves can become self-sustaining—and thus appear in the ear canal as spontaneous emissions—when the total round-trip power gain matches the energy losses (e.g., from viscous damping and acoustic radiation into the ear canal) experienced en route.

The standing-wave model differs fundamentally from the local-oscillator scenario. Rather than supposing that the ''oscillating elements'' generating SOAEs are localized to particular cells or subcellular structures within the organ of Corti, the standing-wave model identifies SOAEs as a *global collective phenomenon* necessarily involving the mechanics, hydrodynamics, and cellular physiology of the entire cochlea, as well as the mechanical and acoustical loads presented to it by the middle and external ears. In the local-oscillator model these macromechanical structures and processes play no fundamental role—they serve merely to connect the autonomous oscillating element with the external environment, providing a conduit for the acoustic energy it produces to escape from the inner ear. In the global standing-wave model, by contrast, the oscillating element comprises the entire cochlea, and the collective response of the hearing organ as a whole contributes essentially to creating, maintaining, and determining the characteristics of the emission.

B. Overview

This paper tests these two alternative models for the origin of mammalian SOAEs. Our discussion focuses on key predictions of the global standing-wave model that distinguish it from the local-oscillator alternative. Although some of these predictions could perhaps be obtained by artful adjustment of local-oscillator models, they all arise quite naturally within the standing-wave framework without the need for additional *ad hoc* assumptions. We therefore begin with a brief review of the global standing-wave model and its distinguishing predictions. In particular, we use the model to derive quantitative answers to questions such as ''What is so special about SOAE frequencies? Why do the SOAEs observed in a particular ear occur at *these* frequencies and not others? and What determines the distribution of SOAE frequency spacings?''[1] Many of the resulting model predictions, which generally involve correlations between spontaneous and evoked emissions and/or the modulation of SOAEs by changes in middle-ear impedance, have been explored in various ways elsewhere (e.g., Kemp, 1979b; Wilson, 1980; Zwicker and Schloth, 1984; Zweig and Shera, 1995; Allen *et al.*, 1995; Talmadge *et al.*, 1998). We extend this earlier work by providing quantitative tests of standing-wave model predictions at frequencies spanning the entire range over which human SOAEs have been reported. Our results provide strong support for the global standing-wave model.

We continue by testing the predictions of two very different versions of the global standing-wave model appearing in the literature. These two alternative standing-wave models differ in the nature of the proposed power source. Whereas the ''passive'' standing-wave model supposes that SOAEs are biological noise, passively amplified by cochlear standing-wave resonances acting as narrow-band nonlinear filters (Allen and Fahey, 1992; Allen *et al.*, 1995; Allen, 2001), the ''active'' standing-wave model supposes that standing-wave amplitudes are actively maintained by coherent wave amplification within the cochlea (Kemp, 1979a; Zweig, 1991; Talmadge and Tubis, 1993; Zweig and Shera, 1995; Talmadge *et al.*, 1998). Our test contributes to the literature on the differences between passive and active SOAE sources (e.g., Bialek and Wit, 1984; Talmadge *et al.*, 1991; Allen *et al.*, 1995; Tubis and Talmadge, 1998; Burns *et al.*, 1998) by providing a compelling empirical demonstration of the special character of SOAEs—amplitude stabilization—that distinguishes them from narrow-band filtered noise. Taken together, our results strongly support the active standing-wave model, which suggests that SOAEs are coherent, amplitude-stabilized acoustic signals produced by the cochlea acting as a biological, hydromechanical analog of a laser oscillator.

II. THE GLOBAL STANDING-WAVE MODEL OF SOAEs

The global standing-wave model of SOAEs posits a close connection between spontaneous otoacoustic emissions and a particular type of evoked emission, namely stimulus-frequency OAEs (or SFOAEs), which are sounds evoked from the ear at the frequency of the stimulus. In a nutshell, the global standing-wave model suggests that SOAEs are

continuously self-evoking stimulus-frequency emissions initiated either by sounds from the environment or by thermal or physiological noise internal to the cochlea.

To deduce characteristics of SOAEs we begin by considering the SFOAE evoked by a low-level pure tone. We interpret the SFOAE as indicating the presence of a backward-traveling wave within the cochlea. To characterize this wave we define the cochlear traveling-wave reflectance, R, as the complex ratio of the out-going (backward-traveling) to the in-going (forward-traveling) pressure wave at the basal end of the cochlea near the stapes (Shera and Zweig, 1993a; Zweig and Shera, 1995; Talmadge *et al.*, 1998):

$$R(f;P_{\text{in}}) \equiv \left.\frac{P_{\text{out}}(f;P_{\text{in}})}{P_{\text{in}}(f)}\right|_{\text{stapes}}. \tag{1}$$

We define $R(f;P_{\text{in}})$ as the reflectance measured at the stapes—rather than introduce a local reflection coefficient (R_x) for every point x within the cochlea (e.g., Kemp, 1979b; Allen, 2001)—because no simple relation between R_x and the total backward-traveling wave exists when wave scattering occurs over a distributed region (Zweig and Shera, 1995).[2] The cochlear reflectance $R(f;P_{\text{in}})$ depends on both the frequency, f, and on the amplitude of the in-going pressure wave, P_{in}. At sound levels in the low-level linear regime near threshold, P_{out} is proportional to P_{in} and R is therefore independent of level (Shera and Zweig, 1993a); at higher sound levels, the amplitude of P_{out} is compressed relative to P_{in} and $|R(f;P_{\text{in}})|$ decreases towards zero. The function $R(f;P_{\text{in}})$ provides a phenomenological characterization of the emission process as seen from the base of the cochlea; according to the coherent-reflection model, its value at any given frequency depends both on the distribution of reflecting impedance perturbations and on the magnitude of round-trip traveling-wave amplitude gains or losses.

By regarding the intervening ear-canal space and middle ear as a linear acousto-mechanical two-port system (Egolf *et al.*, 1992; Peake *et al.*, 1992; Shera and Zweig, 1992b; Puria, 2003) characterized in the frequency domain using reflectance and transmittance coefficients (Shera and Zweig, 1992a; Keefe *et al.*, 1993; Voss and Allen, 1994), one can relate the stimulus-frequency emission measured in the ear canal (P_{SFOAE}) to the value of the cochlear reflectance:

$$P_{\text{SFOAE}} = P_0 G_{\text{mert}} \frac{R(1+R_{\text{stapes}})}{1-RR_{\text{stapes}}}, \tag{2}$$

where $P_0(f)$ is the stimulus source pressure,[3] $G_{\text{mert}}(f)$ is the round-trip middle-ear pressure transfer function,[4] and $R_{\text{stapes}}(f)$ is the reflection coefficient for retrograde cochlear waves at the stapes (Shera and Zweig, 1991a; Zweig and Shera, 1995; Talmadge *et al.*, 1998).[5] The value of R_{stapes} depends not only on the mechanics of the middle ear but also on the acoustic load present in the ear canal (e.g., the ear-canal radiation impedance or the equivalent acoustic impedance of any measurement system placed in the meatus).[6] Note that the values of G_{mert} and R_{stapes} are not independent; energy conservation implies that $|G_{\text{mert}}| \to 0$ as $|R_{\text{stapes}}| \to 1$.

A. Standing waves from multiple internal reflection

The factor $R/(1-RR_{\text{stapes}})$ in Eq. (2) arises from multiple internal reflection within the cochlea, a phenomenon well documented in both the time and frequency domains (e.g., Norton and Neely, 1987; Shera and Zweig, 1993a; Konard-Martin *et al.*, 2001; Dhar *et al.*, 2002). To see this, note that for $|RR_{\text{stapes}}|<1$ the factor is simply the sum of the infinite series

$$R[1+RR_{\text{stapes}}+(RR_{\text{stapes}})^2+\cdots] = R\sum_{n=0}^{\infty}(RR_{\text{stapes}})^n. \tag{3}$$

The terms in this power series can be understood physically as follows. Suppose that the initial forward-traveling wave has unit amplitude at the stapes ($P_{\text{in}}=1$). This primary forward-traveling wave propagates down the cochlear spiral and is partially reflected (re-emitted) in more apical regions of the cochlea. When it returns to the stapes the resulting backward-traveling wave has an amplitude, R, given by the first ($n=0$) term in the power series [see Eq. (1) with $P_{\text{in}}=1$]. At the stapes, the backward wave of amplitude R is then partially reflected back into the cochlea, creating a secondary forward-traveling wave of amplitude RR_{stapes}. This secondary forward-traveling wave is in turn reflected/re-emitted within the cochlea, creating then a secondary backward-traveling wave whose amplitude at the stapes, R^2R_{stapes}, is given by the second ($n=1$) term in the series. The process of multiple reflection continues, each subsequent stapes reflection and cochlear re-emission contributing an additional backward-traveling wave whose amplitude at the stapes differs by a factor of RR_{stapes} from the one before. Adding up all the backward-traveling waves (i.e., summing the power series) yields the factor $R/(1-RR_{\text{stapes}})$ in Eq. (2) for P_{SFOAE}.

Whenever the product RR_{stapes} is positive real, the secondary, tertiary, and all higher-order forward-traveling waves combine in phase with the primary traveling wave at the stapes. The multiple internal reflections then reinforce one another, creating a significant standing-wave component in the cochlear response whose amplitude depends on the product of cochlear and stapes reflection factors, RR_{stapes}. Equation (2) predicts that the standing wave grows without bound as RR_{stapes} approaches 1.[7] In the real cochlea, of course, unconstrained growth is prevented by compressive nonlinearities that limit the energy produced. Once initiated—whether by sounds from the environment or by physiological noise—standing waves of this sort require no external sound for their maintenance;[8] they would be manifest in the ear canal as spontaneous otoacoustic emissions. The standing-wave model thus emphasizes the global nature of SOAEs, with a key role played by the impedance mismatch at the cochlear boundary with the middle ear. Indeed, standing-wave SOAEs would never arise if the stapes and adjoining structures combined to present a perfectly reflectionless boundary ($R_{\text{stapes}} \to 0$).

III. TESTING THE GLOBAL STANDING-WAVE MODEL

Whenever the round-trip amplitude (or standing-wave) gain is sufficient to maintain the emission [i.e., $|RR_{\text{stapes}}|$

$=1$ in the linear analysis of Eq. (2)], the global standing-wave model predicts that SOAEs occur at frequencies f_{SOAE} where RR_{stapes} is positive real. If $\theta(f)$ represents the angle of RR_{stapes}, so that

$$\theta(f) \equiv \angle\{RR_{\mathrm{stapes}}\}, \tag{4}$$

then SOAE frequencies f_{SOAE} satisfy the equation

$$\theta(f_{\mathrm{SOAE}}) = 2\pi n, \tag{5}$$

for some integral value of n. Although neither R nor R_{stapes} is directly accessible noninvasively, a number of indirect tests of the model can be made by using SFOAEs to determine the frequency dependence of $\theta(f)$.

A. The frequency dependence of $\theta(f)$

We obtain the principal frequency dependence of $\theta(f)$ by writing it in the form

$$\theta(f) = \angle R(f) + \angle R_{\mathrm{stapes}}(f), \tag{6}$$

and exploiting two well-established characteristics of SFOAEs at low sound levels.

(1) First, although relative SFOAE amplitudes (i.e., $|P_{\mathrm{SFOAE}}/P_0|$) decrease rapidly with increasing sound intensity, SFOAE phase varies much less strongly with level (e.g., Kemp and Chum, 1980; Zwicker and Schloth, 1984; Shera and Zweig, 1993a). We can therefore estimate $\angle R$ at the relatively low sound levels typical of SOAEs (where $|R|$ is of order 1) using measurements of P_{SFOAE} made at higher levels (where $|R| \ll 1$). [We later correct for the small systematic error introduced by this approximation (see Sec. V A).] When $|R|$ is small Eq. (2) reduces to

$$P_{\mathrm{SFOAE}} \approx P_0 G_{\mathrm{mert}} R(1 + R_{\mathrm{stapes}}) \quad (|R| \ll 1). \tag{7}$$

Solving this equation for R and using the result in Eq. (6) yields

$$\theta(f) \approx \angle P_{\mathrm{SFOAE}} + \angle\{R_{\mathrm{stapes}}/G_{\mathrm{mert}}(1 + R_{\mathrm{stapes}})\}, \tag{8}$$

where $\angle P_0$ has been defined to be zero.

(2) Second, SFOAE phase varies much more rapidly with frequency than do the phases of middle-ear transfer functions. Figure 1 shows a polar plot of typical measurements of $P_{\mathrm{SFOAE}}(f)$ in humans. Although $|P_{\mathrm{SFOAE}}|$ generally varies relatively slowly with frequency, $\angle P_{\mathrm{SFOAE}}(f)$ rotates rapidly, encircling the origin more than 8 times over the frequency range of the data (roughly 1–2 kHz). Since the "angular velocity" (or phase slope) of the rotation is large and nearly constant over intervals comparable to several orbital periods, SFOAE phase changes almost uniformly with frequency. At these frequencies the average orbital period is about 125 Hz; over the frequency range 1–10 kHz, $\angle P_{\mathrm{SFOAE}}(f)$ circles the origin roughly 40 times, yielding an average period of about 225 Hz (Shera and Guinan, 2003). The phases of middle-ear transfer functions, by contrast, vary much more slowly (e.g., Puria *et al.*, 1997; Puria, 2003). Computing $R_{\mathrm{stapes}}/G_{\mathrm{mert}}(1 + R_{\mathrm{stapes}})$ using Puria's (2003) measurements in human cadavers indicates that the middle-ear component of $\theta(f)$ [i.e., the second term in Eq. (8)] amounts to only a single cycle from 1–10 kHz. The middle-ear thus contributes an average phase slope some 40 times smaller than that of P_{SFOAE}.

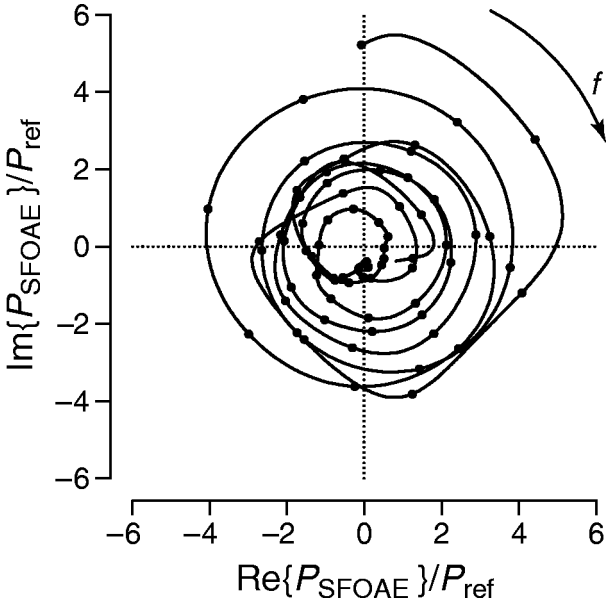


FIG. 1. Polar plot of typical human SFOAEs. The measurements (●) of $P_{\mathrm{SFOAE}}(f)$ were obtained using a variant of the suppression method (Shera and Guinan, 1999) at a stimulus level of 40 dB SPL (subject MAB-L). The axes give the real and imaginary parts of $P_{\mathrm{SFOAE}}/P_{\mathrm{ref}}$, where P_{ref} is 20 μPa. The measurement noise floor is approximately 0.05 in these units. The smooth solid line connecting the data points was computed using bandlimited interpolation. The figure shows $P_{\mathrm{SFOAE}}(f)$ circling clockwise about the origin as frequency increases; the phase traverses more than 8 cycles over the frequency range of the data (0.9–1.9 kHz). Figure 9 of Shera and Guinan (1999) shows these same data in an alternate form (amplitude and phase versus frequency).

Combining items (1) and (2) above, we conclude that as approximated by Eq. (8) the overall frequency dependence of $\theta(f)$ is dominated by $\angle P_{\mathrm{SFOAE}}(f)$. To a good first approximation, Eq. (8) becomes

$$\theta(f) \approx \angle P_{\mathrm{SFOAE}}(f) + \text{constant}, \tag{9}$$

meaning that relative to $\angle P_{\mathrm{SFOAE}}$ the second, "constant" term in Eq. (8) changes slowly with f, at least in humans. We can therefore test the global standing-wave model noninvasively using stimulus-frequency otoacoustic emissions.

IV. QUALITATIVE PREDICTIONS OF THE MODEL

Since $\theta(f)$ evidently rotates through many cycles over the frequency range of human hearing (see Fig. 1), the global standing-wave model predicts that multiple solutions to Eq. (5), and therefore multiple SOAEs, are possible in a single ear. Multiple SOAEs are, in fact, commonly observed experimentally. The study by Talmadge *et al.* (1993) found that roughly 80% of all emitting ears had more than one SOAE (with a median of 5 SOAEs per emitting ear). Since $\theta(f)$

changes (nearly) monotonically, multiple SOAEs can be indexed by the value of n in Eq. (5) and their frequencies written $f_{\mathrm{SOAE}}^{(n)}$.

In the global standing-wave model, the observation that $\theta(f)$ rotates almost uniformly imposes an approximate quantization on the frequency spacing between multiple emissions. The standing-wave quantization condition, namely[9]

$$\theta(f_{\mathrm{SOAE}}^{(n)}) = 2\pi n \quad (n = 0, \pm 1, \ldots), \tag{10}$$

follows from the requirement that forward-traveling waves reflected from the stapes combine in phase with one another. Equation (10) implies that if the round-trip amplitude gain were everywhere sufficient, the spacing between adjacent emissions, Δf_{SOAE}, would be equal to the interval, Δf_{θ}, over which $\theta(f)$ changes by one cycle:

$$\Delta f_{\mathrm{SOAE}}(f) = \Delta f_{\theta}(f). \tag{11}$$

If $\theta(f)$ were to rotate at a constant rate, multiple SOAEs would appear with perfectly regular spacing. But, because the rotation rate of $\theta(f)$ varies on frequency scales both large and small, actual emission spacings never manifest such crystalline regularity. Deviations from a constant rotation rate due to the secular phase curvature of $\theta(f)$ become significant over frequency intervals larger than several orbital periods; as a result, the intervals $\Delta f_{\theta}(f)$ and $\Delta f_{\mathrm{SOAE}}(f)$ vary systematically with f. Over smaller frequency intervals, local variability in $\Delta f_{\theta}(f)$ introduces a more stochastic component to SOAE spacings. In addition to the variability in SOAE spacing arising from $\theta(f)$, spatial fluctuations in the round-trip amplitude gain (i.e., the value of $|RR_{\mathrm{stapes}}|$) produce frequent ''drop-outs'' so that most ears manifest only a handful of the potential SOAEs enumerated by Eq. (10).

Despite these sources of variability, the angle $\theta(f)$ does vary almost uniformly over frequency intervals corresponding to several orbital periods. Locally, $\Delta f_{\theta}(f)$ therefore has a well-defined mean, $\overline{\Delta f}_{\theta}(f)$, whose value depends on frequency. The global standing-wave model then predicts that multiple SOAEs will appear with a corresponding characteristic minimum frequency separation, $\overline{\Delta f}_{\mathrm{SOAE}}(f) = \overline{\Delta f}_{\theta}(f)$, corresponding to $|\Delta n| = 1$. Illustrated in Fig. 2 using a histogram of interemission spacings computed from emission data in the literature (Talmadge *et al.*, 1993; Burns *et al.*, 1992), the existence of a characteristic minimum spacing for SOAEs is well documented (e.g., Schloth, 1983; Dalmayr, 1985; Zwicker, 1988; Russell, 1992; Talmadge *et al.*, 1993). Note that the figure represents SOAE spacings in the fractional form $N_{\mathrm{SOAE}} \equiv \overline{f}_{\mathrm{SOAE}}/\Delta f_{\mathrm{SOAE}}$, where $\overline{f}_{\mathrm{SOAE}}$ is the geometric mean frequency of the two SOAEs. The peak at $N_{\mathrm{SOAE}} \approx 15$ corresponds to a characteristic minimum spacing $\overline{\Delta f}_{\mathrm{SOAE}}$ of approximately 100 Hz for SOAEs near 1500 Hz.

We characterize SOAE spacings using the dimensionless representation N_{SOAE} because previous reports suggest that the characteristic spacing $\overline{\Delta f}_{\mathrm{SOAE}}(f)$ increases in direct proportion to emission frequency, with its value corresponding to a constant fraction of an octave [or to what is essentially the same thing, a constant distance along the basilar membrane or a constant fraction of the psychophysical critical band measured in Barks (e.g., Zwicker, 1989; Talmadge *et al.*, 1993; Braun, 1997)]. It is therefore conventional to express the interval $\overline{\Delta f}_{\mathrm{SOAE}}(f)$ in some relative form (e.g., as a fraction of an octave) rather than directly in Hz. We later demonstrate the systematic deviations from these conventional representations predicted by the global standing-wave model (see Secs. V A and VII E).

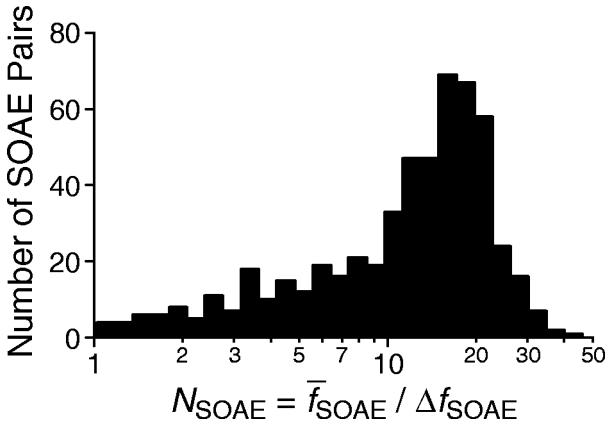


FIG. 2. Histogram of human SOAE spacings. The figure shows a histogram of values of N_{SOAE}, defined as $\overline{f}_{\mathrm{SOAE}}/\Delta f_{\mathrm{SOAE}}$, computed from adjacent pairs of SOAEs reported in compilations of adult (Talmadge *et al.*, 1993) and child (Burns *et al.*, 1992) SOAE data. The spacing Δf_{SOAE} is the absolute value of the difference between the two SOAE frequencies; $\overline{f}_{\mathrm{SOAE}}$ is their geometric mean. The adult data represent 503 SOAE pairs from 67 ears in 44 subjects (age 7–49 years); the child data represent 53 SOAE pairs from 3 children (age 2). Constant values of N_{SOAE} correspond to constant fractions of an octave, where the fraction, r, is given by $r = \log_2(1 + 1/N_{\mathrm{SOAE}})$. For values $N_{\mathrm{SOAE}} \gg 1$, this simplifies to $1/r \approx \ln 2 \cdot N_{\mathrm{SOAE}}$. For $N_{\mathrm{SOAE}} \approx 15$, the spacing is approximately $\frac{1}{10}$ of an octave.

Figure 2 further demonstrates that $\overline{\Delta f}_{\mathrm{SOAE}}$, the characteristic minimum spacing, also represents the most *common* emission spacing. This result can be understood using the global standing-wave model and the fact that the magnitude of RR_{stapes} changes relatively slowly with frequency compared to its angle, $\theta(f)$. As discussed above, both middle-ear reflectances and SFOAE amplitudes typically vary on frequency scales at least several times larger than $\overline{\Delta f}_{\mathrm{SOAE}}$. Thus, if the round-trip amplitude gain is sufficient to stabilize a standing wave at frequency $f_{\mathrm{SOAE}}^{(n)}$, chances are good that the gain will also suffice at the nearby frequencies $f_{\mathrm{SOAE}}^{(n \pm 1)}$.

In the global standing-wave model, SOAE frequencies are determined in part by the impedance mismatch at the cochlear boundary with the middle ear [see Eq. (5)]. Manipulations that modify this basal boundary condition can therefore modulate both SOAE amplitude (by changing $|R_{\mathrm{stapes}}|$ and/or reverse middle-ear transmission) and, more tellingly, SOAE frequency (by changing $\angle R_{\mathrm{stapes}}$). In accord with these predictions, middle-ear impedance changes—as induced, for example, by varying static ear-canal pressure to tension the tympanic membrane or by modifying the impedance of the annular ligament via postural changes that affect static intracochlear fluid pressure—have been found to alter SOAE characteristics, including frequency (e.g., Kemp, 1981; Wilson and Sutton, 1981; Zurek, 1981; Schloth

and Zwicker, 1983; Bell, 1992; Burns *et al.*, 1993; Hauser *et al.*, 1993; de Kleine *et al.*, 2000).

V. QUANTITATIVE PREDICTIONS OF THE MODEL

We can perform more quantitative tests of the global standing-wave model by expressing the quantization condition in the form

$$\left|\frac{d\theta}{df}\right|\Delta f_\theta \approx 2\pi, \tag{12}$$

where we have approximated the phase slope as constant over the interval Δf_θ, consistent with nearly uniform rotation. Using the observation that the frequency dependence of $\theta(f)$ is dominated by $\angle P_{\mathrm{SFOAE}}$ [Eq. (9)], we obtain the estimate

$$\frac{d\theta}{df} \approx \frac{d\angle P_{\mathrm{SFOAE}}}{df} = -2\pi\tau_{\mathrm{SFOAE}}(f), \tag{13}$$

where $\tau_{\mathrm{SFOAE}}(f)$ is the SFOAE phase-gradient (or group) delay. Using this estimate in Eq. (12) yields the prediction $\Delta f_{\mathrm{SOAE}}(f) \approx 1/\tau_{\mathrm{SFOAE}}(f)$, or, equivalently

$$N_{\mathrm{SOAE}}(f) \approx N_{\mathrm{SFOAE}}(f), \tag{14}$$

where $N_{\mathrm{SOAE}} \equiv \overline{f}_{\mathrm{SOAE}}/\Delta f_{\mathrm{SOAE}}$ and $N_{\mathrm{SFOAE}} = f\tau_{\mathrm{SFOAE}}$. The function $N_{\mathrm{SFOAE}}(f)$ is simply SFOAE group delay expressed in periods of the stimulus frequency (Shera and Guinan, 2003). As in Fig. 2, the frequency $\overline{f}_{\mathrm{SOAE}}$ is the geometric mean frequency of the pair of adjacent SOAEs. Equation (14) relates the frequency spacing of *spontaneous* OAEs to the group delay of *evoked* stimulus-frequency OAEs.

A. The characteristic spacing and its dependence on frequency

Figure 3 replots the histogrammed SOAE data (Fig. 2) as a scatterplot versus emission frequency.[10] The density of points is greatest in the upper part of the plot, corresponding to SOAEs separated by intervals close to the characteristic minimum spacing (e.g., $|\Delta n| = 1$). SOAEs separated by wider intervals (e.g., those corresponding to $|\Delta n| > 1$) contribute to the more diffuse appearance in the bottom half. The solid line shows a robust power-law fit, $\overline{N}_{\mathrm{SOAE}}(f)$, to the mode of the distribution. The power-law form of $\overline{N}_{\mathrm{SOAE}}(f)$ appears as a straight line on these log-log axes.

We test Eq. (14) over a four-octave frequency range in the human ear by overlaying the function $\overline{N}_{\mathrm{SFOAE}}(f)$ representing a power-law fit to measurements of human SFOAE group delay (Shera and Guinan, 2003). Agreement between $\overline{N}_{\mathrm{SOAE}}(f)$ and $\overline{N}_{\mathrm{SFOAE}}(f)$ is generally good, although the two lines are somewhat offset from one another, indicating that $1/\overline{\tau}_{\mathrm{SFOAE}}(f)$ slightly overestimates the mean characteristic spacing $\overline{\Delta f}_{\mathrm{SOAE}}(f)$. The parameters of the power-law fits quantify these conclusions (see Table I). Given the estimated uncertainties (95% confidence intervals), the offset between the lines appears significant (compare $\overline{N}_{\mathrm{SOAE}} = 13.7 \pm 0.7$ at 1 kHz versus $\overline{N}_{\mathrm{SFOAE}} = 11.1 \pm 1.2$); the power-law exponents, however, are essentially indistinguishable (compare $\alpha = 0.31 \pm 0.05$ for $\overline{N}_{\mathrm{SOAE}}$ versus $\alpha = 0.37 \pm 0.07$ for $\overline{N}_{\mathrm{SFOAE}}$).

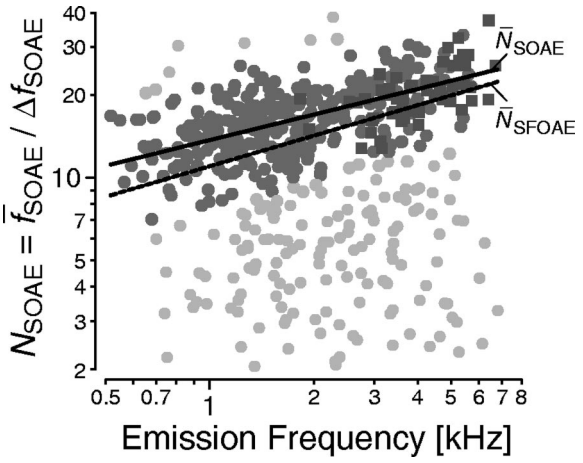


FIG. 3. SOAE spacings compared with SFOAE group delay. The SOAE data from Fig. 2 are presented as a scatterplot showing N_{SOAE} versus emission frequency, $\overline{f}_{\mathrm{SOAE}}$. The circles and squares denote the adult and infant data, respectively. The solid line shows a power-law fit, $\overline{N}_{\mathrm{SOAE}}(f)$, to the peak of the distribution (parameters in Table I). To reduce bias in the fit caused by SOAE pairs with values N_{SOAE} off the peak of the distribution (e.g., data points corresponding to values $|\Delta n| > 1$), a robust loess trend line (Cleveland, 1993) was first computed to locate the approximate mode of the distribution (the results agreed closely with a line drawn by eye). Only data points lying within two standard deviations about the trend line were then included in the fit. Points excluded from the fit are shown in light gray. For comparison, the dashed line shows the power-law fit, $\overline{N}_{\mathrm{SFOAE}}(f)$, computed from measurements of human SFOAE group delay in 9 subjects (Shera and Guinan, 2003).

1. Origins of the discrepancy

Systematic offsets between $\overline{N}_{\mathrm{SOAE}}$ and $\overline{N}_{\mathrm{SFOAE}}$ (see Fig. 3) are expected on theoretical grounds. Recall that the second term in Eq. (8) for $\theta(f)$, approximated as constant in Eq. (9), actually varies slowly with f. The two angles $\theta(f)$ and $\angle P_{\mathrm{SFOAE}}$—the first responsible for quantization of SOAE spacings, the second for SFOAE group delay—therefore rotate at somewhat different rates. The global standing-wave model thus predicts that SOAE frequency spacings generally differ somewhat from the value

TABLE I. Parameters of power-law fits to the functions $N_{\mathrm{SOAE}}(f)$ and $N_{\mathrm{SFOAE}}(f)$. The parameters $\{\alpha,\beta\}$ characterizing the frequency dependence of $N_{\mathrm{SOAE}}(f)$ and $N_{\mathrm{SFOAE}}(f)$ were determined by linear regression using power-law fits of the form $y = \beta x^\alpha$, where y is the dependent variable and $x = f/[\mathrm{kHz}]$ (i.e., frequency or characteristic frequency in kHz). The numbers in parentheses give the approximate uncertainty (i.e., 95% confidence interval) in the final digit(s) estimated from the fits [e.g., $0.31(5) = 0.31 \pm 0.05$]; when the uncertainty is 1 or greater, the position of the decimal point is shown for clarity. The uncertainties in α and $\log\beta$ are strongly correlated, with a typical correlation coefficient between them of roughly -0.8. The parameter β for $\overline{N}^*_{\mathrm{SFOAE}}(f)$ includes a rough correction for the difference in effective sound intensity (see text); the corresponding confidence interval does not include contributions from the appreciable uncertainty in the mean magnitude of round-trip middle-ear transfer functions.

	$\overline{N}_{\mathrm{SOAE}}$	$\overline{N}_{\mathrm{SFOAE}}$	$\overline{N}^*_{\mathrm{SFOAE}}$
α	0.31(5)	0.37(7)	0.37(7)
β	13.7(7)	11.1(1.2)	14.1(1.5)

$1/\overline{\tau}_{\mathrm{SFOAE}}(f)$ given by Eq. (14) (Talmadge *et al.*, 1998). Although the magnitude and sign of this difference depend on middle-ear and transducer characteristics not known with any certainty, Puria's (2003) measurements suggest that it is small (see Sec. III A), amounting to only a fraction of a period at the frequencies explored here.

Differences in the effective intracochlear sound intensities characterizing the two data sets can also produce a systematic offset between $\overline{N}_{\mathrm{SOAE}}$ and $\overline{N}_{\mathrm{SFOAE}}$. Like those shown in Fig. 1, the SFOAE group-delay measurements in Fig. 3 were made at an ear-canal sound level (40 dB SPL) sufficient to ensure that $|R| \ll 1$. SOAE amplitudes, however, are typically somewhat less than 0 dB SPL (e.g., Talmadge *et al.*, 1993). Although measurements of middle-ear transfer functions show considerable intersubject variability, they suggest that these ear-canal SOAE levels result from intracochlear sound pressures roughly the equivalent of those produced by a 15–30-dB SPL stimulus tone (Puria, 2003).[11] Since evoked-emission group delays decrease systematically with increasing sound level (Neely *et al.*, 1988), measurements of $N_{\mathrm{SFOAE}}(f)$ at 40 dB SPL are likely to underestimate values at lower, more comparable sound levels. Although the magnitude of the resulting offset is difficult to estimate with any certainty, taking the effective intracochlear sound-level difference to be roughly 15 dB yields the ''intensity-corrected'' estimate $\overline{N}^{*}_{\mathrm{SFOAE}} = 14.1 \pm 1.5$ at 1 kHz,[12] in closer agreement with the value for $\overline{N}_{\mathrm{SOAE}}$. Note that the reported confidence interval for $\overline{N}^{*}_{\mathrm{SFOAE}}$ does not include contributions from the appreciable uncertainty in the mean magnitude of round-trip middle-ear transfer functions. Despite the considerable uncertainties, this account of the discrepancy between $\overline{N}_{\mathrm{SOAE}}$ and $\overline{N}_{\mathrm{SFOAE}}$ is supported by the analysis of Zweig and Shera (1995), who found no significant difference between the mean value of Δf_{SOAE} averaged over the interval 1–2 kHz and the mean value of $1/\tau_{\mathrm{SFOAE}}$ obtained from 1–2-kHz measurements of ear-canal pressure at a stimulus level of 10 dB SPL.

2. SOAE spacings in other mammals

Although data characterizing the relation between SOAEs and evoked emissions in nonhuman mammals are extremely limited, the available evidence supports the strong correlation between $\overline{N}_{\mathrm{SOAE}}(f)$ and $\overline{N}_{\mathrm{SFOAE}}(f)$ demonstrated here in human ears (Fig. 3). OAE measurements in chinchillas, for example, yield a good correspondence between SOAE spacings and the frequency spacing characteristic of distortion-product (DPOAE) fine structure (Long *et al.*, 2000). Since theoretical and experimental work has demonstrated that DPOAE fine-structure spacing measured at fixed f_2/f_1 is largely determined by SFOAE phase (Talmadge *et al.*, 1998; Kalluri and Shera, 2001), a strong correlation between DPOAE fine structure and SOAE spacings is predicted by the global standing-wave model. Unfortunately, the fact that SOAEs detectable in chinchillas generally occur at higher frequencies than in humans complicates a direct interspecies comparison. Extrapolating the power-law form of the human $\overline{N}_{\mathrm{SOAE}}(f)$ to the chinchilla modal SOAE frequency of 10 kHz suggests that chinchilla SOAE spacings are roughly a factor of 2 larger than their counterparts in humans. This difference is consistent [cf. Eq. (14)] with the observation that human SFOAE group delays are longer, by roughly a factor of 2–3, than those in laboratory animals such as cats and guinea pigs (Shera and Guinan, 2003; Zwicker and Manley, 1981).

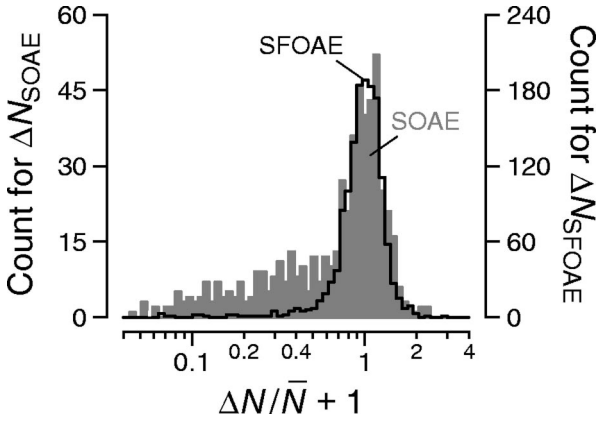


FIG. 4. Variability of SOAE spacings and SFOAE group delay. The figure shows histograms of the deviations $\Delta N_{\mathrm{SOAE}}/\overline{N}_{\mathrm{SOAE}}$ (shaded gray) and $\Delta N_{\mathrm{SFOAE}}/\overline{N}_{\mathrm{SFOAE}}$ (black line), where $\Delta N \equiv N - \overline{N}$. The deviations $\Delta N/\overline{N}$ have been pooled across frequency and are plotted on a logarithmic scale by showing the value $\Delta N/\overline{N}+1$ along the abscissa. For ΔN_{SOAE}, residuals about the power-law fit, $\overline{N}_{\mathrm{SOAE}}(f)$, were computed from the data shown in Fig. 3; the histogram is based on 556 measurements in 47 subjects. For $\Delta N_{\mathrm{SFOAE}}$, residuals about $\overline{N}_{\mathrm{SFOAE}}(f)$ were computed from the data of Shera and Guinan (2003); the histogram is based on 1441 measurements in 9 subjects. The vertical scales (left and right axes) differ by a factor of 4, a value set by the ratio of histogram areas on the interval $1 \pm \sigma$, where $\sigma \approx 0.25$ is the standard deviation of the SFOAE histogram about its peak. SOAEs spaced at wide intervals corresponding to $|\Delta n| > 1$ contribute to the long tail of the SOAE histogram that extends leftwards along the abscissa.

B. Variability of SOAE spacings

Figure 4 demonstrates that the global standing-wave model also accounts for the evident variability in the spacing of adjacent SOAEs (see Figs. 2 and 3). The figure shows histograms of the deviations $\Delta N_{\mathrm{SOAE}}/\overline{N}_{\mathrm{SOAE}}$ and $\Delta N_{\mathrm{SFOAE}}/\overline{N}_{\mathrm{SFOAE}}$, where in each case $\Delta N \equiv N - \overline{N}$. About their peaks the two distributions are nearly identical. They differ in their tails because the SOAE histogram includes data from adjacent SOAEs with wide frequency spacings. According to the global standing-wave model, SOAEs at wide spacings result from gaps in the series created by ''dropouts.'' Because of variations across frequency in the value of $|RR_{\mathrm{stapes}}|$, not all frequencies $f^{(n)}_{\mathrm{SOAE}}$ in Eq. (10) are realized as SOAEs and SOAEs separated by intervals corresponding to $|\Delta n| > 1$ often occur. The SOAE histogram is therefore skewed leftward in the tail. Despite this expected difference in the tails of the two distributions, the strong quantitative match between the peaks supports the global standing-wave model, which predicts that the variability in close SOAE spacings results from variations in $\tau_{\mathrm{SFOAE}}(f)$. The theory of coherent reflection filtering traces this variability in $\tau_{\mathrm{SFOAE}}(f)$ to the randomness of the underlying impedance perturbations filtered by properties of the traveling wave (see Sec. VII D).

C. Frequency shifts due to changes in middle-ear impedance

The standing-wave quantization condition [Eq. (5)] can be used to predict the magnitude and sign of the SOAE frequency shifts induced by changes in middle-ear impedance. To obtain the relation, we let the load impedance presented to the cochlea by the middle and external ears depend on some parameter whose unperturbed value we denote by κ. For example, κ might be the stiffness of the annular ligament or the static pressure in the ear canal. For an SOAE of frequency f the standing-wave quantization condition implies that

$$\angle R(f)+\angle R_{\text{stapes}}(f,\kappa)=2\pi n, \tag{15}$$

where κ appears among the independent variables that determine R_{stapes}. Imagine now that we modify the middle-ear load by taking $\kappa\rightarrow\kappa+\delta\kappa$. Because of the resulting change in stapes reflectance, the standing-wave quantization condition [Eq. (15)] is no longer satisfied at frequency f. To maintain the standing wave, the SOAE frequency must shift ($f\rightarrow f+\delta f$). When the equality in Eq. (15) has been restored

$$\angle R(f+\delta f)+\angle R_{\text{stapes}}(f+\delta f,\kappa+\delta\kappa)=2\pi n, \tag{16}$$

where we have assumed that all changes are small enough that n remains invariant. Equations (15) and (16) imply that the net phase change due to the combined effects of $\delta\kappa$ and δf must be zero. To first order in $\delta\kappa$ and δf

$$(\partial_f\angle R+\partial_f\angle R_{\text{stapes}})\,\delta f+(\partial_\kappa\angle R_{\text{stapes}})\,\delta\kappa=0, \tag{17}$$

where we have used the notational shorthand $\partial_x\equiv\partial/\partial x$. Since $\angle R_{\text{stapes}}$ rotates much less rapidly with frequency than $\angle R$ (i.e., $|\partial_f\angle R_{\text{stapes}}|\ll|\partial_f\angle R|$; see Sec. III A), we obtain the relation

$$\delta f\approx-\left(\frac{\partial_\kappa\angle R_{\text{stapes}}}{\partial_f\angle R}\right)\delta\kappa. \tag{18}$$

Equation (18) can be put in a form more convenient for comparison with experiment by writing it in terms of fractional changes:

$$\delta f/f\approx\left(\frac{\partial_{\ln\kappa}\angle R_{\text{stapes}}}{2\pi N_{\text{SFOAE}}}\right)\delta\kappa/\kappa, \tag{19}$$

where we have used $N_{\text{SFOAE}}\equiv f\cdot\tau_{\text{SFOAE}}$ with $\tau_{\text{SFOAE}}\approx-\partial_f\angle R/2\pi$. Equation (19) relates fractional changes in SOAE frequency to fractional changes in the middle-ear parameter κ. Note that the result is completely general and does not depend on any particular form for R_{stapes}.

1. Frequency shifts due to middle-ear stiffness changes

We can use Eq. (19) to predict the relative signs of δf and $\delta\kappa$ under manipulations that change the effective stiffness of the middle ear. We proceed by combining Eq. (19) with Puria's (2003) measurements and model of $R_{\text{stapes}}(f)$ in human temporal bones. Puria has shown that the general trends in his measurements of $R_{\text{stapes}}(f)$ are well captured by a simple model that approximates the cochlear input impedance as resistive and the middle-ear load seen from the cochlea by a series combination of spring, mass, and damper.[13]

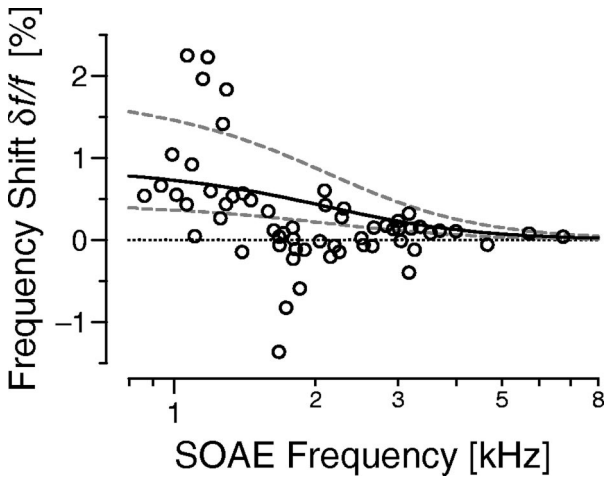


FIG. 5. Comparison of measured and predicted effect of postural changes on SOAE frequencies. Measured SOAE frequency shifts (○, 60 SOAEs from 13 ears) from the study of de Kleine *et al.* (2000) are shown together with standing-wave-model predictions based on Eq. (19), with κ taken to be the effective stiffness of the middle-ear load. The quantity $\partial_{\ln\kappa}\angle R_{\text{stapes}}$ was evaluated using Puria's (2003) simple model of $R_{\text{stapes}}(f)$; $N_{\text{SFOAE}}(f)$ was approximated by the power-law fit shown in Fig. 3. The solid line corresponds to a fractional stiffness increase $\delta\kappa/\kappa$ of 50%; the dashed lines to increases of 25% and 100% (lower and upper curves, respectively). The dotted line marks the location of the zero along the ordinate.

Using this approximation and taking κ to be the spring constant of the middle-ear load yields $\partial_{\ln\kappa}\angle R_{\text{stapes}}>0$. According to Eq. (19), δf and $\delta\kappa$ therefore have the same sign. In other words, the global standing-wave model predicts that increasing the stiffness of the middle-ear system (e.g., by tensing the eardrum or stretching the annular ligament) generally increases SOAE frequencies, in agreement with the changes observed experimentally (e.g., Kemp, 1981; Wilson and Sutton, 1981; Zurek, 1981; Schloth and Zwicker, 1983; Hauser *et al.*, 1993; de Kleine *et al.*, 2000).

2. Magnitude, sign, and frequency dependence of the shifts

Figure 5 quantifies these remarks by comparing standing-wave-model predictions with the posture-induced changes in SOAE frequencies measured by de Kleine *et al.* (2000). Equation (19) was evaluated using Puria's (2003) model of $R_{\text{stapes}}(f)$ and the power-law fit to measurements of $N_{\text{SFOAE}}(f)$ shown in Fig. 3 (Shera and Guinan, 2003). Since the magnitudes of posture-induced changes in the stiffness of the middle-ear load are not known with any certainty, and presumably vary from subject to subject depending on such things as the patency of the cochlear aqueduct, we show predictions using three different values of $\delta\kappa/\kappa$ corresponding to increases of 25%, 50%, and 100%. Estimated stiffness increases of roughly this magnitude were obtained in studies of middle-ear transmission in which the impedance of the annular ligament was varied using postural shifts (Avan *et al.*, 2000; Büki *et al.*, 2000). These estimates are also broadly consistent with the results of Pang and Peake (1986), who made direct measurements of the stiffness of the annular ligament during stapedius-muscle contractions in cat and found increases ranging up to a factor of 10; changes induced by postural shifts are likely to be significantly smaller.

Although the measured SOAE frequency shifts show considerable scatter, Fig. 5 demonstrates that the global standing-wave model [Eq. (19)] quantitatively accounts for the major features reported in the data (e.g., de Kleine *et al.*, 2000). These features include the typical magnitude of the SOAE frequency shifts (1–2%), their most common direction (toward higher frequencies), and their overall variation with SOAE frequency (largest below 2 kHz). The scatter in the data presumably results from intersubject variation and details of the middle-ear load not captured by the simple model of $R_{\text{stapes}}(f)$. For example, Puria's measurements indicate that although $\angle R_{\text{stapes}}(f)$ generally decreases over the frequency range shown in Fig. 5, the change is not perfectly smooth. Unlike the model, the data manifest considerable local variation in the magnitude and sign of the phase slope. These local departures from the overall trend are presumably responsible for much of the observed variation in SOAE frequency shifts (e.g., the decrease in SOAE frequencies near 2 kHz).

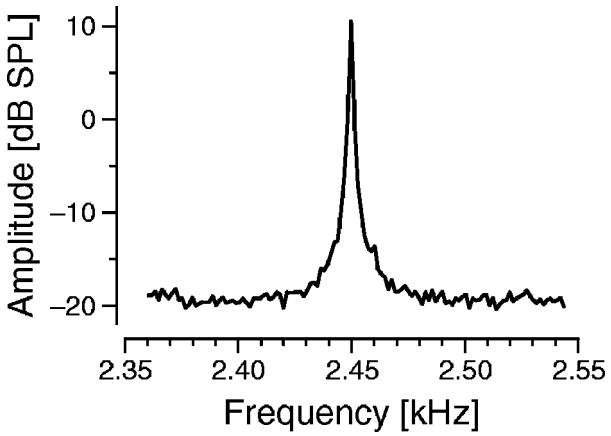


FIG. 6. Frequency spectrum for a typical human SOAE. The figure shows the spectrum obtained by averaging the spectral amplitudes of 89 contiguous segments of SOAE waveform, each of approximately 671 ms duration (subject WL-R). The total averaging time was about 60 seconds. Measurement methods are described in footnote 14.

3. Middle-ear contributions to emission bandwidths

Although we have focused here on artificially induced changes in middle-ear impedance, continuous small perturbations in the basal boundary condition presumably occur naturally from a number of sources, including variations in middle-ear cavity pressures due to breathing, swallowing, or blood flow; variations in the stiffness of the annular ligament due to spontaneous stapedius contractions; and changes in intracochlear pressure related to heartbeat. According to the global standing-wave model, these and other ongoing physiological perturbations produce small corresponding fluctuations in SOAE frequencies. Equation (19) relating $\delta f/f$ to changes $\delta\kappa/\kappa$, where κ can be any parameter that modifies $\angle R_{\text{stapes}}$, predicts that fractional SOAE frequency shifts are inversely proportional to N_{SFOAE}. In other words, the longer the SFOAE group delay [i.e., the more rapidly $\angle R(f)$ rotates with frequency], the more stable are SOAE frequencies against perturbations in the boundary conditions. All other things being equal, the global standing-wave model thus predicts that species with long SFOAE group delays should generally have more stable SOAEs (i.e., narrower SOAE bandwidths) than species with shorter delays. Interestingly, this is precisely the trend observed experimentally (e.g., Ohyama *et al.*, 1991; Long *et al.*, 2000): SOAEs in guinea pigs and chinchillas, species with relatively short group delays, are generally less stable and have broader bandwidths than SOAEs in humans, where SFOAE group delays are substantially longer (Shera and Guinan, 2003).

VI. ARE THE STANDING WAVES POWERED BY NOISE?

Most global standing-wave models in the literature propose that mammalian SOAEs—such as the one whose spectrum is illustrated in Fig. 6—result from intracochlear standing waves stabilized by a balance between round-trip energy losses and level-dependent coherent wave amplification. Nevertheless, standing-wave resonances driven solely by thermal or other noise sources would also appear as narrow-band acoustic signals in the ear canal (Allen and Fahey, 1992; Allen, 2001). If the standing-wave resonances were of sufficiently high Q, the ear-canal spectral characteristics of the SOAEs generated by a ''passive,'' noise-driven standing-wave model would be indistinguishable from the amplitude-stabilized coherent signals produced by the ''active'' standing-wave model.

A. Generalization of the Bialek–Wit histogram

Although they cannot be distinguished by their spectra, the acoustic signals generated by the two different models (i.e., narrow-band filtered noise versus amplitude-stabilized oscillations) can be distinguished by the statistical properties of their time waveforms, as Bialek and Wit (1984) were the first to point out for SOAEs. Bialek and Wit showed that the SOAE pressure waveform has a double-humped, non-Gaussian probability distribution inconsistent with the output of a linear passive narrow-band filter driven by noise. Talmadge *et al.* (1991) subsequently showed that double-humped distributions rule out many nonlinear passive systems as well. Here, we extend the Bialek–Wit analysis to explicitly demonstrate the amplitude stabilization that distinguishes SOAEs from narrow-band filtered noise. The Bialek–Wit histogram is a one-dimensional projection of the more general two-dimensional distribution described here.

To obtain the distribution, we begin by writing the time-varying ear-canal pressure due to an SOAE in the form

$$p_{\text{SOAE}}(t) = p(t)\cos[2\pi f_{\text{SOAE}} t + \phi(t)], \tag{20}$$

where f_{SOAE} is the nominal SOAE frequency and $p(t)$ and $\phi(t)$ are its slowly varying amplitude and phase. We then represent the SOAE waveform by the complex phasor

$$\tilde{p}_{\text{SOAE}}(t) \equiv p(t)e^{i\phi(t)}. \tag{21}$$

Over time, the phasor $\tilde{p}_{\text{SOAE}}(t)$ moves about in the complex plane, tracing out a trajectory whose instantaneous polar coordinates (i.e., radial distance from the origin and angle with the real axis) are $p(t)$ and $\phi(t)$, respectively. The real and imaginary parts of $\tilde{p}_{\text{SOAE}}(t)$ are thus the in-phase and

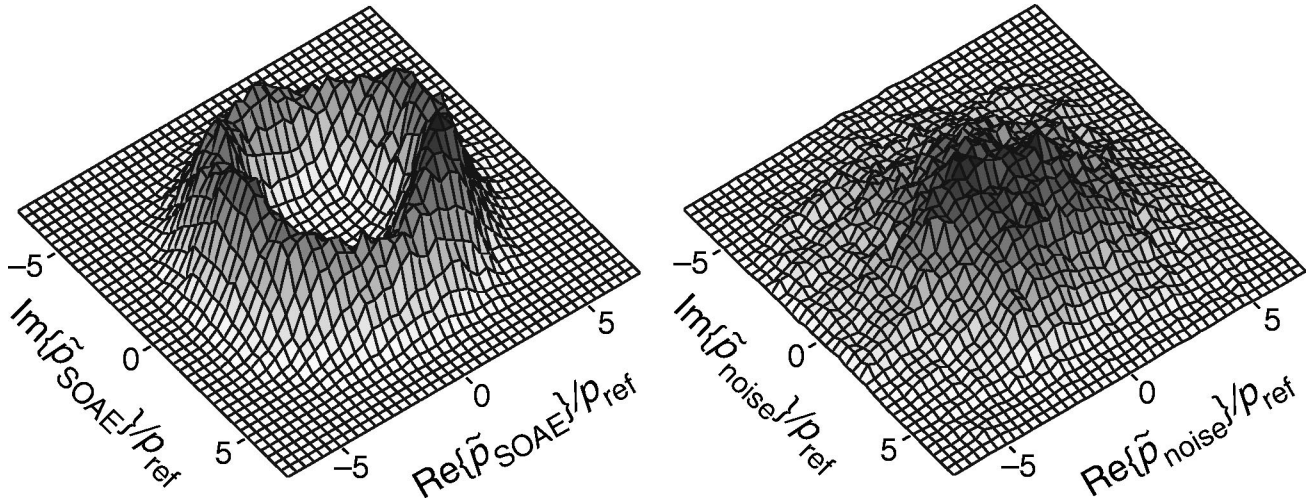


FIG. 7. Probability distributions $\mathcal{D}[\tilde{p}_{\mathrm{SOAE}}(t)]$ and $\mathcal{D}[\tilde{p}_{\mathrm{noise}}(t)]$ for an SOAE and for a narrow-band noise signal with the same power spectrum. Distributions were computed from 80 s of data as described in footnote 14 for the human SOAE shown in Fig. 6 (left-hand panel) and a noise signal with the identical power spectrum (right-hand panel). The x- and y axes represent the real and imaginary parts (in-phase and quadrature components) of the complex phasors $\tilde{p}(t)$ described in the text. Pressures are shown in units of $p_{\mathrm{ref}}=20\ \mu$Pa.

quadrature components obtained by heterodyning the SOAE waveform using a reference signal of frequency f_{SOAE}.

B. Molehills and moleruns

Figure 7 (left-hand panel) gives the probability distribution $\mathcal{D}_{\mathrm{S}}\equiv\mathcal{D}[\tilde{p}_{\mathrm{SOAE}}(t)]$ obtained from the SOAE shown in Fig. 6.[14] For comparison, and as a control, the right-hand panel gives the corresponding distribution, $\mathcal{D}_{\mathrm{N}}\equiv\mathcal{D}[\tilde{p}_{\mathrm{noise}}(t)]$, for a noise signal filtered in such a way that its power spectrum is identical to that of the SOAE. [The noise signal was obtained by passing random-phase, flat-spectrum noise through a filter with frequency response equal to the amplitude spectrum of the SOAE, as illustrated in Fig. 6. The phasor $\tilde{p}_{\mathrm{noise}}(t)$ was subsequently computed as described for $\tilde{p}_{\mathrm{SOAE}}(t)$ in footnote 14.] Both distributions are shown as functions of the complex variable $\tilde{p}$. In other words, the x and y axes give the possible amplitudes of the $\cos(2\pi f_{\mathrm{SOAE}}t)$ and $\sin(2\pi f_{\mathrm{SOAE}}t)$ components of the signal. The height of the surface above any small element of area in the plane is proportional to the long-term average probability that the in-phase and quadrature components of the signal, as measured during any particular short interval of time, will be found to lie within the specified range.

For the filtered-noise signal, the phasor $\tilde{p}_{\mathrm{noise}}(t)$ wanders randomly about in the complex plane at a rate inversely proportional to the signal bandwidth. This random motion appears superimposed on a circular drift whose angular velocity is proportional to the difference between the instantaneous signal frequency and the center frequency of the spectrum. The resulting probability distribution, $\mathcal{D}_{\mathrm{N}}$, is that of a two-dimensional Gaussian ''molehill'' centered on the origin (Rice, 1954).

As illustrated in Fig. 7, the statistics of the SOAE signal are strikingly different from filtered noise. Rather than resembling a molehill, the probability distribution suggests a ''molerun'' or volcanic caldera. Near the origin, where $\mathcal{D}_{\mathrm{N}}$ is largest, $\mathcal{D}_{\mathrm{S}}$ is vanishingly small. Significant values of $\mathcal{D}_{\mathrm{S}}$ are confined to a relatively narrow ring of finite radius. Evidently, the SOAE amplitude is stabilized at a nonzero value about which it manifests only small fluctuations (e.g., due to thermal noise and other physiological perturbations). SOAE phase, by contrast, exhibits no such stabilization; over time, the phase wanders through all possible angles, its random motion again superimposed on a circular drift whose speed and direction change erratically according to the magnitude and sign of transient deviations in SOAE frequency about its central value.[15] Although the absolute phase reference necessary to stabilize SOAE phase is not available physiologically, phase stabilization can be achieved by entraining the emission to an external tone.

Projecting the caldera-like distribution $\mathcal{D}_{\mathrm{S}}$ shown in the left-hand panel of Fig. 7 onto the xz plane (i.e., computing the distribution $\mathcal{D}[\mathrm{Re}\{\tilde{p}_{\mathrm{SOAE}}(t)\}]$) yields a double-humped, non-Gaussian histogram similar to those reported previously in mammals, lizards, and birds (e.g., Bialek and Wit, 1984; Talmadge *et al.*, 1991; van Dijk *et al.*, 1996).[16] Unlike the Bialek–Wit distribution, however, the caldera $\mathcal{D}_{\mathrm{S}}$ clearly demonstrates that fluctuations in SOAE amplitude are limited both from above and from below [see also Fig. 3 of Talmadge *et al.* (1991)]. SOAEs amplitudes are evidently stabilized within a narrow range and, in particular, almost never fall appreciably below their mean value. Although amplitude stabilization of this sort can readily be understood in the context of the active standing-wave model (see Sec. VII A), it appears difficult if not impossible to reconcile with the passive standing-wave model and its assumption that SOAEs are powered entirely by randomly fluctuating, incoherent biological noise.

VII. DISCUSSION

Our quantitative tests provide strong support for the global standing-wave model and its prediction that SOAE frequencies are determined by RR_{stapes} (e.g., Kemp, 1979a, b; Zweig and Shera, 1995; Talmadge *et al.*, 1998). The results demonstrate that in addition to predicting the existence of multiple emissions with a characteristic minimum frequency

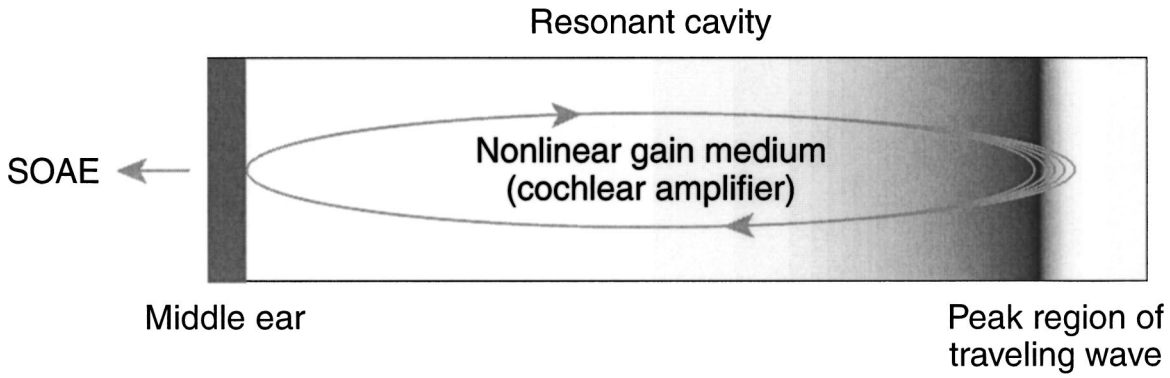


FIG. 8. Analogy between the cochlear production of SOAEs and the coherent emission of light by an optical laser. The region between the stapes and the peak of the traveling wave is represented as a resonant cavity enclosing a nonlinear gain medium. The gradient in shading illustrates that most of the amplification occurs just basal to the peak of the traveling wave. Partial reflection of forward- and backward-traveling waves occurs at each end of the cavity. At the apical end, coherent reflection occurs over a distributed region spanning the peak of the traveling wave. Standing waves occur at frequencies for which the round-trip phase change is an integral multiple of 2π. Standing-wave amplitudes are stabilized when the round-trip gain matches the round-trip losses due to internal damping and acoustic emission through the middle ear into the environment.

spacing, the global standing-wave model also accurately predicts the mean value of this spacing, its standard deviation, and its power-law dependence on SOAE frequency. Furthermore, the global standing-wave model accounts for the magnitude, sign, and frequency dependence of changes in SOAE frequency that result from modulations in middle-ear stiffness. Although some of the SOAE characteristics tested and accounted for here might be replicable in artfully constructed local-oscillator models (e.g., by *ad hoc* adjustment of SOAE spacings), they all arise quite naturally and immediately in the global standing-wave framework. Finally, the statistics of SOAE time waveforms demonstrate that SOAEs are coherent, amplitude-stabilized signals. Taken together, these tests of the global standing-wave model support the proposed mechanism-based classification of SOAEs within the group of reflection-source OAEs (Shera and Guinan, 1999). In addition, they provide compelling evidence that mammalian SOAEs constitute a global collective phenomenon—amplitude-stabilized cochlear standing waves—rather than the local, autonomous oscillation of some cellular constituent of the organ of Corti.

A. Analogy with a laser oscillator

The existence of amplitude-stabilized standing waves within the cochlea suggests that the cochlea is acting as a biological, hydromechanical analog of a laser oscillator (see also Zweig, 1991; Russell and Kössl, 1999; Kemp, 2002). Reduced to its essentials, a laser oscillator consists of a resonant cavity enclosing a gain medium that supports coherent wave amplification (see Fig. 8). In the cochlea, the ''resonant cavity'' spans the region between the stapes and the peak of the traveling wave. At either end of this region, cochlear traveling waves are partially reflected back into the cavity. At the stapes, backward-traveling waves reflect due to the impedance mismatch with the middle ear; at the apical end of the cavity, forward-traveling waves undergo coherent reflection near the peak of the traveling wave. With partially reflecting ''mirrors'' at both ends, the cochlea differs from a typical optical laser, in which one of the mirrors is made perfectly reflecting. In further contrast with an optical laser, where reflection back into the cavity occurs at a well-defined location (i.e., in the plane of the reflecting mirror), the coherent reflection near the peak of the traveling wave is distributed over a finite region of the cochlea (i.e., the width of the traveling-wave envelope). The cochlear gain medium consists of the cellular force generators and surrounding mechanical, hydrodynamic, and electrical processes known collectively as the ''cochlear amplifier.'' In contrast to an optical laser, where amplification occurs uniformly throughout the cavity, amplification in the cochlea occurs almost entirely near the apical mirror (i.e., in the region just basal to the peak of the traveling wave).

On each pass through the cavity, waves are amplified by their interaction with the gain medium. At certain special frequencies—those for which the round-trip phase change is an integral multiple of 2π—multiple internal reflection creates standing waves. If the round-trip gain matches the round-trip losses (e.g., due to damping and acoustic radiation into the environment), stable oscillations can result that appear in the ear canal as SOAEs. Just as in an optical laser, oscillation amplitudes are self-stabilizing. Since the cochlear amplifier is limited in the energy it can produce, the gain medium is nonlinear, with the amount of amplification decreasing as the wave amplitude grows. Standing-wave amplitudes are therefore stable against perturbations: If some random fluctuation increases the wave amplitude slightly, the round-trip gain decreases a little and the wave amplitude falls back down. Conversely, if the wave amplitude decreases, the total gain increases, pulling the oscillation amplitude back up. Amplitude stabilization of this sort produces SOAE signals with the statistical properties illustrated in Fig. 7. Similar probability distributions characterize the coherent radiation generated by optical lasers (e.g., Golay, 1961; Siegman, 1986).

Unlike optical lasers, the cochlea can emit at multiple, nonharmonically related frequencies. In an optical laser, the cavity size is fixed and tuned to a single frequency and its harmonics. Wave propagation in the cochlea, however, is highly dispersive and the location of the wave peak—and hence the location of the partially reflecting mirror—depends on frequency. Consequently, the round-trip phase condition

is satisfied at many frequencies and the cochlea can produce multiple SOAEs simultaneously.

B. Sources of initial traveling-wave energy

The global standing-wave model indicates that the multiple internal reflection and amplification of traveling-wave energy responsible for SOAEs can be initiated simply by sounds from the environment and/or by physiological noise. In addition, autonomous oscillations of cells or subcellular structures may sometimes kick-start the emission process by acting as initial generators of traveling-wave energy. For example, should some ''resonant element'' within the cochlea suddenly begin to oscillate mechanically, the resulting movements presumably create backward-traveling waves (if the oscillation frequency is less than or equal to the local CF). These backward-traveling waves subsequently reflect off the stapes, thereby initiating the process of multiple internal reflection characteristic of global standing-wave resonances.

Even though the original disturbance may arise through the action of some local oscillator, long-term stability requires that any final SOAE frequency be consistent with Eq. (5), so that round-trip phase shifts equal an integral multiple of 2π. Unless the initiating oscillator can adapt by changing its frequency to satisfy this global constraint (or is largely impervious to the perturbing influence of its own energy, which is fed back to it by reflection off the stapes),[17] the oscillator's output will be highly unstable and unlikely to persist. Computational studies of ''completely active'' cochlear models consisting of an array of coupled van der Pol oscillators confirm the importance of cochlear standing waves and the middle-ear boundary condition in determining SOAE frequencies (van Hengel *et al.*, 1996).

C. Relationship to phenomenological oscillator models

Properties of SOAEs such as their interactions with one another and with external tones have been successfully described by representing individual SOAEs using a nonlinear, limit-cycle oscillator such as the van der Pol (e.g., Murphy *et al.*, 1995a, b, 1996; van Dijk and Wit, 1990a, b). These phenomenological, limit-cycle oscillator models were not, as a rule, developed to describe the ''oscillating elements'' within the cochlea; rather, their aim was to approximate the behavior of a complex system of equations (such as those describing the generation of standing waves in an active, nonlinear transmission line) by a single effective oscillator, thereby providing simple, analytically tractable representations of SOAEs as they appear in the ear canal. Thus, these phenomenological models are not inconsistent with the global standing-wave model, which proposes a mechanism by which SOAEs originate within the cochlea. The evident success of the global standing-wave model contradicts the notion, often implicit in the local-oscillator framework, that SOAEs measured in the ear canal provide direct access to the local elementary cellular oscillators within the organ of Corti (e.g., Sisto and Moleti, 1999).

D. The mechanism of reflection within the cochlea

According to the global standing-wave model, SOAEs are simply a special case of a more general and ubiquitous otoacoustic phenomenon: the production of reflection-source OAEs by the ear. Although many predictions of the model depend only on the empirical form of the cochlear traveling-wave reflectance, $R(f)$, understanding the mechanism by which reflection-source OAEs (such as low-level stimulus-frequency and transient-evoked emissions) originate provides deeper insight.

Considerable evidence suggests that the generation of reflection-source OAEs is well described by the theory of coherent reflection filtering (Shera and Zweig, 1993b; Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 2003). The theory indicates that reflection-source OAEs arise via coherent reflection from densely and randomly distributed impedance perturbations. These perturbations presumably include both those clearly visible in the anatomy, such as spatial variations in outer-hair-cell (OHC) number and geometry (e.g., Engström *et al.*, 1966; Bredberg, 1968; Wright, 1984; Lonsbury-Martin *et al.*, 1988), as well as morphologically less conspicuous perturbations, such as variations in OHC forces due to random, cell-to-cell variations in hair-bundle stiffness or the number of somatic motor proteins.

The coherent-reflection model predicts that the SFOAE evoked by a tone comprises a sum of wavelets scattered by perturbations located throughout the peak of the traveling wave. The SFOAE therefore arises from a distributed region, roughly equal in extent to the width of the traveling-wave envelope. In the 1–2-kHz region of the human cochlea, this distance spans on the order of 100 rows of outer hair cells at sound levels near threshold (Zweig and Shera, 1995). Although Kemp's (1979b) original standing-wave model postulated that the backward-traveling wave originates from a point reflection, the coherent-reflection model indicates that this wave necessarily arises over a region equivalent to the span of many hair cells.[18]

The theory of coherent-reflection filtering resolves other problems with the original standing-wave model. For example, Kemp (1979b) pointed out that an ''apparently random dependence [of the amplitude of the reflectance] on frequency for each ear ... is necessary to account for the presence of strong resonances at some frequencies and not others.'' The coherent-reflection model explains this random variation in reflectance magnitude by representing the total backward wave as the sum of many wavelets having an irregular distribution of amplitudes and phases dependent on the particular array of impedance perturbations encountered within the peak of the traveling wave. This same mechanism also produces random variations in reflectance phase over frequency, giving rise to the distribution of SOAE spacings about their mean value (Fig. 4).

According to the coherent-reflection theory, quantitative features of the distribution of SOAE spacings such as its central value and relative spread are determined dynamically by properties of the traveling wave (Zweig and Shera, 1995; Talmadge *et al.*, 1998). For example, the characteristic mini-

mum spacing at frequency f is set by the value of the wavelength, $\hat{\lambda}$, near the peak of the wave envelope:

$$\overline{N}_{\mathrm{SOAE}}(f) \approx \overline{N}_{\mathrm{SFOAE}}(f) \approx 2N_{\mathrm{BM}}(f), \tag{22}$$

where

$$N_{\mathrm{BM}} = l/\hat{\lambda}, \tag{23}$$

and is a function of location (or characteristic frequency) within the cochlea.[19] In these equations the parameter l represents the distance over which the characteristic frequency (CF) changes by a factor of e, and the factor of 2 originates in round-trip phase shifts and the Bragg scattering condition. According to the theory, the observed increase in the value of $\overline{N}_{\mathrm{SOAE}}$ towards the base of the cochlea (Fig. 3) reflects the systematic decrease in the wavelength $\hat{\lambda}$, which diminishes at a rate of roughly 25% per octave with increasing CF (Shera and Guinan, 2003).

Since Kemp's (1979b) original standing-wave model did not include the effects of traveling-wave propagation gains and/or losses,[20] the model needed to associate large reflection coefficients with many points along the basilar membrane (at a minimum, those corresponding to SOAE frequencies) in order to generate sizable standing waves. In the coherent-reflection model, by contrast, cochlear traveling waves are amplified as they propagate toward and away from the site of scattering; values of $|R|>1$ measured at the stapes are therefore readily obtained with only small perturbations in the mechanics (Talmadge and Tubis, 1993; Zweig and Shera, 1995). Since small, densely distributed perturbations can produce large values of $|R|$, the modern standing-wave model predicts that most SOAEs result from normal mechanical variability rather than from pathologically large impedance discontinuities. [A possible exception may be the relatively rare class of ''atypical SOAEs'' characterized by their unusually large amplitudes and frequent association with significant audiometric abnormality (reviewed in Lonsbury-Martin and Martin, 2001).]

In contrast to the local-oscillator model, which typically emphasizes the pathology of the condition by ascribing SOAEs to a ''disturbance'' or outright ''failure'' of some local feedback control mechanism, the global standing-wave model emphasizes the *normality* of most SOAEs by demonstrating both their close relation to other types of evoked OAEs (i.e., low-level SFOAEs and other reflection-source emissions) and their origin as the expected consequence of distributed wave amplification and reflection in the presence of small, nonpathological impedance perturbations. In this way, the global standing-wave model resolves the paradox noted by Geisler (1998) in his discussion of the van der Pol oscillator as a local-oscillator model for SOAEs:

> ''[Why] doesn't every section of the cochlea act that way [i.e., as a limit-cycle oscillator] and [the cochlea therefore] produce emissions at all frequencies? It follows that there must be something different about those cochlear sites that generate the relatively few emissions observed. Unfortunately, the search for such differences has not been successful.''

Viewed from the perspective of the global standing-wave model, the failure of this search for differences is not surprising. Indeed, cochlear sites corresponding to SOAE frequencies need manifest no special distinguishing features. In the global standing-wave model, SOAE frequencies are determined by RR_{stapes}, and SOAEs therefore trace their origin to aspects of the mechanics as subtle—and as *nonlocal* to the site in question—as the magnitude and angle of the impedance mismatch at the cochlear boundary with the middle ear, the spatial-frequency content of the cochlear impedance perturbations that scatter the wave, and the total round-trip traveling-wave gain and phase shift experienced en route.

E. Representations of the characteristic minimum spacing

SOAE frequency spacings are conventionally displayed using a histogram that masks any frequency dependence in the data (e.g., Fig. 2). As a result, the characteristic minimum spacing, $\overline{\Delta f}_{\mathrm{SOAE}}$, has been somewhat misrepresented in the literature. We note, for example, that the frequency dependence of $\overline{N}_{\mathrm{SOAE}}(\overline{f}_{\mathrm{SOAE}})$ evident in the scatterplot of Fig. 3 demonstrates that Δf_{SOAE} corresponds neither to a constant fraction of an octave, to a constant distance along the basilar membrane, nor to a constant fraction of the psychophysical critical band measured in Barks (e.g., Zwicker, 1989; Russell, 1992; Talmadge *et al.*, 1993; Braun, 1997).

Although the human characteristic spacing $\overline{\Delta f}_{\mathrm{SOAE}}(f)$ bears no simple relationship to Zwicker's critical band rate scale [contrary to folklore in the field, $\overline{\Delta f}_{\mathrm{SOAE}}(f) \neq 0.4$ Bark], the spacing $\overline{\Delta f}_{\mathrm{SOAE}}(f)$ does roughly correspond to a constant fraction of the equivalent rectangular bandwidth (ERB) estimated from otoacoustic and behavioral measurements at frequencies greater than 1 kHz (Shera *et al.*, 2002; Oxenham and Shera, 2003). Using the power-law approximations to $\overline{\Delta f}_{\mathrm{SOAE}}(f)$ and $\mathrm{ERB}(f)$ [Table I and Shera *et al.* (2002)] yields $\overline{\Delta f}_{\mathrm{SOAE}}(f) \approx \beta(f/[\mathrm{kHz}])^{\alpha} \cdot \mathrm{ERB}(f)$, where $\beta = 0.93 \pm 0.08$ and $\alpha = -0.01 \pm 0.08$. Note that the power-law exponent is indistinguishable from zero. The approximate proportionality between $\overline{\Delta f}_{\mathrm{SOAE}}$ and the psychophysical ERB demonstrated here should not be taken to imply some direct causal connection between human cochlear tuning and SOAE spacings. According to the global standing-wave model and the theory of coherent reflection filtering, the proportionality arises because both SFOAE phase slopes (and hence SOAE spacings) and the bandwidths of peripheral auditory filters are, for very different reasons, strongly correlated to the group delays of basilar-membrane transfer functions (Shera and Guinan, 2003).

We emphasize that $\overline{\Delta f}_{\mathrm{SOAE}}$ is merely the mode of a distribution of spacings and therefore has no absolute significance; inappropriate reification of its value can lead to dubious conclusions. Braun (1997), for example, argues against the global standing-wave model in order to propose an ostensible connection, mediated via efferent feedback from the inferior colliculus, between SOAE spacings and psychophysical critical bands. Braun proceeds by plotting the distribution of spacings computed from all possible SOAE pairs (i.e., an all-order distribution rather than the distribution of

first-order intervals computed from adjacent SOAEs as in Fig. 2) The all-order distribution shows only weak or nonexistent peaks at multiples of the characteristic minimum spacing $\overline{\Delta f}_{\mathrm{SOAE}}$ (which he calls the ''preferred minimum distance,'' or PMD). Braun concludes that the ''total lack of any system of multiples of the PMD in the interval distribution curve contradicts all concepts that assume a spectral periodicity of SOAEs.'' Braun's argument, however, neglects both the frequency dependence of $\overline{\Delta f}_{\mathrm{SOAE}}(f)$ and the substantial variability in the distribution of first-order intervals at fixed frequency (see Fig. 3). Because adjacent intervals are largely independent of one another, the nonzero width of the first-order distribution (see Fig. 4) smears out the distribution of higher-order intervals. For example, the distribution of second-order intervals can be approximated by convolving the first-order distribution with itself and is therefore broader than the first-order distribution by roughly a factor of 2. As a result of this smearing, peaks corresponding to higher-order intervals are difficult or impossible to discern in the all-order histogram.

1. The corresponding length scale

The spatial distance along the basilar membrane that corresponds to the modal frequency spacing $\overline{\Delta f}_{\mathrm{SOAE}}(f_{\mathrm{SOAE}})$ can be found as a function of SOAE frequency by noting that exponential position–frequency functions map constant distances Δx to constant relative frequency intervals $\Delta f/f$. According to Table I, $\overline{N}_{\mathrm{SOAE}}(\overline{f}_{\mathrm{SOAE}})=\overline{f}_{\mathrm{SOAE}}/\overline{\Delta f}_{\mathrm{SOAE}}$ varies with emission frequency as $\overline{f}^{\alpha}_{\mathrm{SOAE}}$ with $\alpha\approx 0.31\pm 0.05$; if the human cochlear map is exponential (Greenwood, 1990), the distance $\overline{\Delta x}_{\mathrm{SOAE}}$ corresponding to $\overline{\Delta f}_{\mathrm{SOAE}}$ therefore varies as $1/\overline{f}^{\alpha}_{\mathrm{SOAE}}$.

We can relate the distance $\overline{\Delta x}_{\mathrm{SOAE}}$ to the wavelength, $\hat{\lambda}$, of the traveling wave at its peak by combining these results with Eqs. (22) and (23). The calculation yields

$$\overline{\Delta x}_{\mathrm{SOAE}}\approx \tfrac{1}{2}\hat{\lambda}. \tag{24}$$

In other words, the characteristic places associated with nearest-neighbor SOAEs are separated, on average, by a distance equal to one-half the wavelength of the traveling wave. According to the coherent-reflection theory, variations about this modal value occur not because $\hat{\lambda}$ varies irregularly with position, but because the distribution of frequency spacings Δf_{SOAE} reflects an underlying disorder in the process that creates backward-traveling waves by reversing the forward flow of traveling-wave energy (i.e., scattering by random impedance perturbations). For clarity, we emphasize that the coherent-reflection theory predicts that the length scale $\overline{\Delta x}_{\mathrm{SOAE}}$ arises dynamically and need have no geometric correlate in the structure or mechanics of the cochlea (Zweig and Shera, 1995).

F. Strategies for testing the standing-wave model

The principal distinguishing prediction of the global standing-wave model is that SOAE frequencies are determined by RR_{stapes}. Specifically, the model predicts that SOAE frequencies f_{SOAE} satisfy the standing-wave quantization-condition [Eq. (5)], which requires that RR_{stapes} be positive real; equivalently

$$\theta(f_{\mathrm{SOAE}})=0 \pmod{2\pi}, \tag{25}$$

where $\theta(f)\equiv\angle\{RR_{\mathrm{stapes}}\}$. Direct experimental confirmation (or refutation) of Eq. (25) would provide a definitive test of the global standing-wave model. Unfortunately, although possible in principle, this test is difficult in practice. Rigorous tests of Eq. (25) require measurement of both $\angle R(f)$ and $\angle R_{\mathrm{stapes}}(f)$ at SOAE frequencies in a single ear. Furthermore, in the case of $\angle R(f)$, the measurement must be made at intracochlear sound intensities equivalent to those normally produced by the SOAEs in question. One is then left in the unfortunate position that the very SOAEs one is trying to understand ineluctably compromise and contaminate the measurements of $\angle R(f)$ needed to test the theory.[21]

Our strategy in this paper has been to recast the model predictions of Eq. (25) in a theoretically less definitive but experimentally more tractable form. Rather than probe the absolute phase of RR_{stapes} at frequencies f_{SOAE}, we examine the distribution of human SOAE spacings by using noninvasive measurements of SFOAEs to estimate the frequency dependence of $\angle R(f)$ and measurements of middle-ear pressure transfer functions (Puria, 2003) to argue that middle-ear contributions to the distribution of spacings Δf_{SOAE} are generally small, at least in human ears. The results provide strong but indirect support for Eq. (25) in humans (see Figs. 2–4). In addition, we use the standing-wave quantization condition to predict the effect of changes in middle-ear boundary conditions on SOAE frequencies. Although incomplete knowledge of both $\angle R_{\mathrm{stapes}}(f)$ and the magnitude of posture-induced stiffness changes in individual ears precludes a more rigorous comparison with experiment, the global standing-wave model successfully reproduces the major trends in the data (Fig. 5).

Despite their inherent limitations, our tests provide strong support for the idea that human SOAEs arise via global standing-wave resonances. Although definitive examination of the standing-wave model awaits a direct experimental probe of Eq. (25), extending the measurements and ideas developed here to SOAEs in other species, mammalian and nonmammalian alike, would provide important tests of the generality of the model.

G. SOAEs in nonmammals

Although the evidence reported here suggests that most mammalian SOAEs are amplitude-stabilized standing waves, local autonomous oscillation of some cellular constituent of the ear may, of course, underlie some subset of mammalian SOAEs (e.g., the ''atypical SOAEs'' mentioned above). The local-oscillator mechanism may also operate in species such as frogs, lizards, and birds, in which spontaneous cellular oscillations have been observed (e.g., Crawford and Fettiplace, 1985; Denk and Webb, 1992; Martin *et al.*, 2001) but which appear to lack basilar-membrane traveling waves (e.g., Peake and Ling, 1980; Manley, 1990).

Even in these species, however, we conjecture that mechanisms analogous to those posited by the global standing-wave model may often be operating. Note, for example, that the frequency selectivity of primary auditory afferent fibers in some bird and lizard species can evidently match or exceed that found in many mammalian cochleae at comparable frequencies (Manley, 2001), and associated with this tuning are substantial frequency-dependent group delays (e.g., Hillery and Narins, 1984; Smolders and Klinke, 1986; Gleich and Narins, 1988; Manley *et al.*, 1990). If frequency tuning in these species arises or is manifest mechanically, one expects significant mechanical group delays—and thus the possibility of global standing-wave-like resonances—despite the apparent absence of a clear analog of the basilar-membrane traveling wave. Although the phase changes associated with frequency tuning appear inextricably linked to basilar-membrane motion in the mammalian cochlea, the global standing-wave model imposes no such requirement. A ''resonant cavity'' containing a nonlinear ''gain medium''—comprising in this context a slightly irregular array of tuned oscillators, all coupled together through the surrounding fluids and/or via ancillary structures (e.g., the tectorial membrane) to produce the large mechanical phase shifts concomitant with sharp tuning—may be all that's required for creating global resonances analogous to those evidently responsible for SOAEs in the mammalian ear.

A common origin in global, standing-wave-like resonances may account for many of the otherwise puzzling similarities among mammalian and nonmammalian SOAEs. For example, in humans and lizards the frequency spacing between multiple SOAEs appears roughly comparable (within a factor of 2 or 3), even though the human organ of Corti is roughly 17–170 times longer than lizard papillae in overall length (Manley, 1990, 2001). The global standing-wave model explains this seemingly paradoxical observation by implying that the length of the hearing organ is essentially irrelevant to the generation of SOAEs. In the global standing-wave model, SOAE spacings are determined not by the cochlear distance between putative ''oscillating elements'' but by the frequency dependence of mechanical phase shifts; that is, by the characteristics of mechanical tuning, a functional arena where the differences between humans and lizards are less pronounced than in the anatomy (Manley, 2001). To understand more clearly the similarities and differences between mammalian and nonmammalian SOAEs, one needs to extend the types of measurements used here (e.g., determination of quantitative relationships between evoked and spontaneous OAEs and their modulation by changes in external boundary conditions) to the nonmammalian ear.

H. Implications for the cellular basis of the cochlear amplifier

The local-oscillator model of SOAEs plays a central role in recent discussions of the cellular basis of the mammalian cochlear amplifier. For example, Hudspeth and colleagues (e.g., Hudspeth, 1997; Martin and Hudspeth, 1999; Eguíluz *et al.*, 2000; Martin and Hudspeth, 2001; Martin *et al.*, 2001, 2003; Jülicher *et al.*, 2003) argue that the cellular constituents of the ear's amplifier operate near a ''critical point'' (specifically, near a so-called Andronov–Hopf bifurcation)[22] where spontaneous cellular oscillation sets in. In this view, dating back to the work of Gold (1948), SOAEs occur when the cellular control mechanisms needed to hold a hair cell close to the critical point break down:

> ''Because self-tuning positions the system slightly on the oscillating side of the critical point, self-tuned criticality provides a natural explanation for [spontaneous] otoacoustic emissions. In its normal working state, the inner ear would generate faint sounds with a broad range of frequencies. If the feedback mechanism were to fail in certain cells, the spontaneous oscillations could become large enough for distinct tones to be emitted.'' (Camalet *et al.*, 2000).

In this guise, the local-oscillator model has been repeatedly invoked to argue that outer-hair-cell (OHC) somatic motility is unlikely to constitute the ''active process'' in mammalian hearing. Martin and Hudspeth (1999, 2001; Martin *et al.*, 2003), for example, question the role of OHC somatic motility in the generation of mammalian SOAEs (see also Köppl, 1995; Martin *et al.*, 2001; Duke, 2002; Jülicher *et al.*, 2003). They note that although spontaneous movements have been reported in nonmammalian hair bundles ''there have been no observations of spontaneous outer-hair-cell contractions.'' Since they claim that ''unprovoked mechanical oscillations of some constituent of the inner ear *must* underlie the production of spontaneous otoacoustic emissions'' (Martin *et al.*, 2003, emphasis added), the evident conclusion is that mammalian SOAEs—and, by implication, the mammalian cochlear amplifier—require active hair-bundle motility.[23]

But Martin and Hudspeth's argument rests on a logical fallacy: that what's true of the whole must be true of the parts (the fallacy of division); that because the ear oscillates spontaneously some of its cellular constituents must do so themselves, and, furthermore, that only the autonomous oscillation of the ear's constituent parts can explain the spontaneous emission of sound by the whole.

We argue here, however, that most mammalian SOAEs arise not through autonomous cellular oscillations but through the collective action of the entire cochlea. SOAEs are not primarily the result of ''unprovoked movements'' of cellular constituents of the cochlear amplifier; rather, they are continually self-evoking stimulus–frequency OAEs that arise via the multiple internal reflection and coherent amplification of traveling-wave energy within the cochlea. As Kemp and others long ago suggested, the ''bifurcation'' responsible for the creation of self-sustaining SOAEs can arise *globally* rather than locally, much like that in a laser oscillator (e.g., Kemp, 1979a, b; Talmadge and Tubis, 1993; Zweig and Shera, 1995).

The global standing-wave model thus reverses the chain of causality inherent in the local-oscillator model: the mammalian ear does not whistle because individual hair cells oscillate spontaneously; rather, hair cells oscillate spontaneously because the ear whistles. Not only is the whole ''more than the sum of its parts,'' but the parts acquire new proper-

ties by virtue of their embedding in the whole. The apparent absence of spontaneous contractions in isolated OHCs has been presented as *primae facie* evidence against a role for somatic motility as the primary motor element underlying the mammalian cochlear amplifier (e.g., Martin and Hudspeth, 1999, 2001; Martin *et al.*, 2001, 2003; Jülicher *et al.*, 2003). Although a strong reductionist bias in our thinking may render these arguments superficially compelling, the logic that supports them is evidently fallacious.

Many qualitative emission phenomena, including the spontaneous production of sound, are no doubt generic to mechanical systems (such as the inner ears of frogs, lizards, birds, and mammals) that contain coupled arrays of active, nonlinear oscillators tuned to different frequencies. Although the existence and qualitative features of SOAEs may be generic, their quantitative properties (e.g., their bandwidths, their frequency separations, their interactions with one another and with external tones, their relationship to other emission types) presumably depend on details of the morphology, the physiology, and the hydrodynamics of the systems that produce them. Indeed, quantitative differences in features such as these do distinguish mammalian and nonmammalian SOAEs (e.g., van Dijk and Wit, 1990a; Köppl, 1995; van Dijk *et al.*, 1998), and to some extent even the SOAEs from different mammals (e.g., Ohyama *et al.*, 1991; Long *et al.*, 2000). Just as one can construct functional lasers around a variety of gain media, so one expects to find qualitative family resemblances among SOAEs arising from hearing organs that employ diverse strategies for enhancing their frequency selectivity and sensitivity to sound (e.g., Martin and Hudspeth, 1999; Liberman *et al.*, 2002); much more surprising indeed would be the finding that a single quantitative theory provides a satisfactory explanation for them all.

ACKNOWLEDGMENTS

The author gratefully acknowledges stimulating discussions with and/or helpful comments on the manuscript from John J. Guinan, Jont B. Allen, Pim van Dijk, Paul F. Fahey, K. Domenica Karavitaki, Geoffrey A. Manley, Stephen T. Neely, Sunil Puria, Carrick L. Talmadge, Arnold Tubis, Robert H. Withnell, and an anonymous reviewer. This work was supported by Grant No. R01 DC03687 from the NIDCD, National Institutes of Health.

[1]Previewing answers to these questions provided by the global standing-wave model, we note that the model predictions hinge on the ''standing-wave quantization condition,'' which determines the set of possible SOAE frequencies via Eq. (5) and characteristic SOAE spacings via Eq. (11). According to the theory of coherent-reflection filtering, the modal value of the characteristic SOAE spacing can be traced to the value of the wavelength of the traveling wave at its peak (or, equivalently, to the group delay at the peak of the basilar-membrane transfer function).

[2]For example, the cochlear reflectance $R(f)$ can be written as the product of R_x with the forward ($T^+_{0\to x}$) and reverse ($T^-_{x\to 0}$) traveling-wave transfer functions (e.g., $R=T^+_{0\to x}R_xT^-_{x\to 0}$) only in the unrealistic case that wave reflection/re-emission at frequency f occurs at the single point x.

[3]The stimulus source pressure $P_0(f)$ is the ear-canal pressure that would be measured if R were zero. In practice it is usually obtained by calibrating the stimulus earphone at moderate to high sound intensities (i.e., by rescaling the earphone pressure measured when $|P_{\mathrm{SFOAE}}/P_0|\ll 1$).

[4]G_{mert} is the product, $T^+_{\mathrm{me}}T^-_{\mathrm{me}}$, of the forward and reverse middle-ear pressure transfer functions measured at sound intensities where $|P_{\mathrm{SFOAE}}/P_0|\ll 1$. The forward pressure transfer function is defined by $T^+_{\mathrm{me}}\equiv(P_{\mathrm{v}}-P_{\mathrm{t}})/P_{\mathrm{ec}}$, where P_{v} and P_{t} are, respectively, the pressures in the scala vestibuli and scala tympani at the base of the cochlea; P_{ec} is the pressure in the ear canal. The middle ear is presumed driven in the ''forward'' direction with a sound source in the ear canal. Similarly, the reverse pressure transfer function is defined by $T^-_{\mathrm{me}}\equiv P_{\mathrm{ec}}/(P_{\mathrm{v}}-P_{\mathrm{t}})$, where the middle ear is now presumed driven from within the cochlea. Note that T^-_{me} depends on the characteristics of the transducer assembly used to measure P_{ec}; Eq. (2) assumes that this same transducer is used during the measurement of P_{SFOAE}. Since $|P_{\mathrm{t}}/P_{\mathrm{v}}|\ll 1$ (Nedzelnitsky, 1980), T^+_{me} and T^-_{me} can be approximated by measurements made in the scala vestibuli (e.g., $T^+_{\mathrm{me}}\approx P_{\mathrm{v}}/P_{\mathrm{ec}}$).

[5]Equation (2) assumes that the cochlear input impedance at frequency f is approximately real at intensities such that $|P_{\mathrm{SFOAE}}/P_0|\ll 1$ (e.g., Aibara *et al.*, 2001). Reality of the cochlear input impedance is guaranteed by the tapering symmetry manifest, for example, by the cochlea of the cat (Shera and Zweig, 1991b). Shera and Zweig (1991a) discuss the more general case in which the cochlear input impedance is complex and the characteristic wave impedances of the cochlear transmission line depend on the direction of propagation.

[6]The reflection coefficient R_{stapes} has the value $R_{\mathrm{stapes}}=r^-+[t^+t^-R_s/(1-r^+R_s)]$, where the functions $r^\pm$ and $t^\pm$ are the reflectance and transmittance coefficients characterizing the residual ear-canal space and middle-ear (Shera and Zweig, 1992a) and R_s is the Thévenin-equivalent reflectance of the ear-canal acoustic load. Talmadge *et al.* (1998) derive formulas for R_{stapes} and G_{mert} valid for specific models for the middle-ear, ear-canal space, and earphone.

[7]The equation $RR_{\mathrm{stapes}}=1$ thus defines the ''instability modes'' of this linear analysis (Talmadge and Tubis, 1993; Talmadge *et al.*, 1998). Note that because little energy escapes from the cochlea as $|R_{\mathrm{staples}}|\to 1$, the boundary conditions optimal for producing large intracochlear standing waves are simultaneously those making their detection in the ear canal most difficult.

[8]In the active model, standing-wave amplitudes are actively maintained by wave amplification; in the passive model, they are driven by ongoing biological noise.

[9]Kemp's (1979b) original standing-wave model uses an approximate form of the quantization condition valid over frequency intervals roughly corresponding to $|\Delta n|\lesssim 2$. Kemp's condition can be obtained from Eq. (5) by approximating $\angle R(f)$ by $-\omega T(f)$, where $T(f)$ is the emission group delay at the stapes, defined by $T(f)=-d\angle R/d\omega$, with $\omega=2\pi f$. Kemp's formula would be equivalent to Eq. (5) if $T(f)$ were independent of f; in fact, however, $T(f)\sim 1/f^{1-\alpha}$ with $\alpha\approx 0.37$ (see Table I and Shera and Guinan, 2003). Kemp's formula applies over frequency intervals sufficiently small that $T(f)$ can be regarded as constant. An estimate, Δf_{K}, of the size of this interval follows from the requirement that $|d\ln T/df|\Delta f_{\mathrm{K}}<\epsilon$, where $\epsilon\ll 1$. Evaluating the logarithmic derivative and expressing Δf_{K} in terms of the characteristic spacing $\overline{\Delta f}_{\mathrm{SOAE}}$ yields $\Delta f_{\mathrm{K}}<\epsilon[\bar{N}_{\mathrm{SOAE}}/(1-\alpha)]\overline{\Delta f}_{\mathrm{SOAE}}$, where $\bar{N}_{\mathrm{SOAE}}$ is the modal value of N_{SOAE}. Taking $\epsilon\sim 0.15$ and $\bar{N}_{\mathrm{SOAE}}\approx 15$ at 1500 Hz yields $\Delta f_{\mathrm{K}}/\overline{\Delta f}_{\mathrm{SOAE}}\lesssim 4$, or, equivalently, $|\Delta n|\lesssim 2$.

[10]SOAEs with frequencies within 2 Hz of the value $2f_1-f_2$ computed by using other SOAE pairs in the same ear as primaries f_1 and f_2 were tagged as likely to represent lower-sideband cubic distortion products formed by the interaction of other, independent SOAEs in the same ear (cf. Burns *et al.*, 1984; Talmadge *et al.*, 1993). SOAE pairs that included one or more of these probable ''distortion-product SOAEs'' (16 of the 503 pairs of adult emissions) were excluded both from Fig. 3 and from the power-law fit.

[11]Comparisons of psychoacoustic and emission cancellation levels for $(2f_1-f_2)$-combination-tones (Zwicker and Harris, 1990) suggest that a 0-dB SPL SOAE may produce as much intracochlear excitation as a 35–50-dB SPL tone presented in the ear canal. Interpretation of these measurements, however, is complicated by a number of factors, including uncontrolled mixing of reflection- and distortion-source DPOAE components (e.g., Kalluri and Shera, 2001) and uncertainties associated with the psychoacoustic cancellation paradigm (e.g., Smoorenburg, 1974; Siegel and Borneman, 1999).

[12]According to the measurements and analysis of Neely *et al.* (1988), evoked OAE latencies measured at sound level $L+\Delta L$ (in dB SPL) will

differ from those measured at level L by a frequency-independent factor of approximately $c^{-\Delta L/100}$, with $c=5$. This value is likely to be an upper bound; Neely *et al.*'s characterization presumably overestimates the intensity dependence of OAE latencies at low sound levels, where cochlear mechanics becomes approximately linear.

[13]R_{stapes} has the form $R_{\text{stapes}}=(Z_{\text{mer}}-Z_{\text{c}})/(Z_{\text{mer}}+Z_{\text{c}})$, where Z_{c} is the cochlear input impedance (assumed resistive) and Z_{mer} is the ''reverse middle-ear impedance'' (Puria, 2003).

[14]With the subject reclining comfortably in a sound-isolated chamber, ear-canal pressure was transduced using an Etymotic Research ER-10c microphone and digitized at a sampling rate of 12.2 kHz using a Tucker-Davis System III running custom software. Eighty seconds of data were filtered using a tenth-order, recursive-exponential bandpass filter (Shera and Zweig, 1993a; Kalluri and Shera, 2001) centered on the nominal SOAE frequency ($f_{\text{SOAE}}=2449.6$ Hz). The filter bandwidth (approximately 200 Hz) was taken to be much larger than the emission bandwidth (approximately 1 Hz). The analytic signal, $z_{\text{SOAE}}(t)$, of the filtered waveform was then obtained using the Hilbert transform (e.g., Cohen, 1995) and the function $\tilde{p}_{\text{SOAE}}(t)$ computed using the equation $\tilde{p}_{\text{SOAE}}(t)=z_{\text{SOAE}}(t)e^{-2\pi i f_{\text{SOAE}}t}$. Multiplication by the phasor $e^{-2\pi i f_{\text{SOAE}}t}$ isolates the slowly varying components of $z_{\text{SOAE}}(t)$ by dividing out the rapid phase rotation due to the carrier (SOAE) frequency. Temporal resolution was improved by a factor of 4 using bandlimited $\sin(x)/x$ interpolation prior to computing the probability distribution $\mathcal{D}[\tilde{p}_{\text{SOAE}}(t)]$.

[15]Many of these dynamic features are best illustrated by animating the time evolution of $\mathcal{D}[\tilde{p}(t)]$ (see http://epl.meei.harvard.edu/~shera/soae-movie.zip). To keep the animation short (and continually interesting) the time interval between movie frames was increased as the square of the elapsed time. The statistical power of the distribution therefore increases linearly from frame to frame.

[16]The Bialek–Wit analysis actually involves computing the distribution $\mathcal{D}[p_{\text{SOAE}}(t)]=\mathcal{D}[\text{Re}\{z_{\text{SOAE}}(t)\}]=\mathcal{D}[\text{Re}\{\tilde{p}_{\text{SOAE}}(t)e^{2\pi i f_{\text{SOAE}}t}\}]$ (see footnote 14 for an explanation of the notation). Because it includes smearing effects due to rapid phase rotation at the carrier frequency (f_{SOAE}), the Bialek–Wit distribution is generally much smoother than $\mathcal{D}[\text{Re}\{\tilde{p}_{\text{SOAE}}(t)\}]$. The two distributions are nearly identical, however, when computed over time intervals long compared to the reciprocal of the emission bandwidth.

[17]For example, an autonomous cellular oscillator would not experience strong, bidirectional coupling to the external environment if the stapes and adjoining structures somehow conspired to present a nearly reflectionless boundary. Although normal, in vivo values of R_{stapes} remain uncertain, typical transducer assemblies have Thévenin-equivalent reflectances close to 1 in magnitude; the condition $|R_{\text{stapes}}|\ll 1$ therefore seems unlikely to apply in most experimental situations.

[18]Incoherent reflection from large punctate perturbations may, however, dominate in certain pathologies or in specialized cochleae, such as in the ''auditory fovea'' of the CF-FM bat (e.g., Kössl and Vater, 1995), in which the mechanical properties of the cochlear partition change rapidly with position.

[19]Local scaling symmetry implies that N_{BM} also equals the group delay of the basilar-membrane transfer function at its peak, measured in periods of the CF (Shera and Guinan, 2003).

[20]Kemp's (1979b) original standing-wave model was published before the term ''cochlear amplifier'' had even been coined (Neely, 1983; Davis, 1983).

[21]A similar problem is encountered when trying to determine whether the middle-ear reflectance has magnitude greater than 1 at frequencies near SOAEs (e.g., Allen *et al.*, 1995; Tubis and Talmadge, 1998; Burns *et al.*, 1998).

[22]For a critique of this claim, see Zweig (2003), and the comments and discussion reprinted following the text.

[23]Even when a possible amplificatory function of OHC somatic motility is acknowledged, SOAEs are invariably presumed to arise via autonomous subcellular oscillations, with the hair bundle retaining its role as putative causal agent (e.g., Duke, 2002): ''In the mammalian cochlea, the outer hair cells are widely believed to power the movement of the basilar membrance. It remains unclear, however, whether the outer-hair-cell motor is itself a Hopf oscillator, or whether it is simply an additional linear amplifier that boosts oscillations generated by the hair bundle. Spontaneous oscillations of an outer hair cell have never been observed.'' Jülicher *et al.* (2003) argue along similar lines: ''The concept of self-tuned critical oscillations applies essentially to all vertebrate ears since they all have spontaneous otoacoustic emissions. However, the question of what are the underlying specific mechanisms at the origin of spontaneous oscillations is a subject of active research. In mammals, it is widely thought that outer hair cells are the active elements and thus might contain mechanical oscillators. However, even though it is well established that outer hair cells can contract their cell bodies in the presence of electrical stimuli, no spontaneous oscillations have so far been observed.''

Aibara, R., Welsh, J. T., Puria, S., and Goode, R. L. (**2001**). ''Human middle-ear sound transfer function and cochlear input impedance,'' Hear. Res. **152**, 100–109.

Allen, J. B., and Fahey, P. F. (**1992**). ''Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane,'' J. Acoust. Soc. Am. **92**, 178–188.

Allen, J. B., Shaw, G., and Kimberley, B. P. (**1995**). ''Characterization of the nonlinear ear canal impedance at low sound levels,'' Assoc. Res. Otolaryngol. Abs. **18**, 797.

Allen, J. B. (**2001**). ''Nonlinear cochlear signal processing,'' in *Physiology of the Ear*, edited by A. F. Jahn and J. Santos-Sacchi (Singular, San Diego), pp. 393–442.

Avan, P., Büki, B., Maat, B., Dordain, M., and Wit, H. P. (**2000**). ''Middle-ear influence on otoacoustic emissions. I. Noninvasive investigation of the human transmission apparatus and comparison with model results,'' Hear. Res. **140**, 189–201.

Bell, A. (**1992**). ''Circadian and menstrual rhythms in frequency variations of spontaneous otoacoustic emissions from human ears,'' Hear. Res. **58**, 91–100.

Bialek, W., and Wit, H. P. (**1984**). ''Quantum limits to oscillator stability: Theory and experiments on acoustic emissions from the human ear,'' Phys. Lett. **104A**, 173–178.

Braun, M. (**1997**). ''Frequency spacing of multiple spontaneous otoacoustic emissions shows relation to critical bands: A large-scale cumulative study,'' Hear. Res. **114**, 197–203.

Bredberg, G. (**1968**). ''Cellular patterns and nerve supply of the human organ of Corti,'' Acta Oto-Laryngol., Suppl. **236**, 1–135.

Büki, B., Chomicki, A., Dordain, M., Lemaire, J., Wit, H. P., Chazal, J., and Avan, P. (**2000**). ''Middle-ear influence on otoacoustic emissions. II. Contributions of posture and intracranial pressure,'' Hear. Res. **140**, 202–211.

Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). ''Prevalence of spontaneous otoacoustic emissions in neonates,'' J. Acoust. Soc. Am. **91**, 1571–1575.

Burns, E. M., Harrison, W. A., Bulen, J. C., and Keefe, D. H. (**1993**). ''Voluntary contraction of middle ear muscles: Effects on input impedance, energy reflectance and spontaneous otoacoustic emissions,'' Hear. Res. **67**, 117–127.

Burns, E. M., Keefe, D. H., and Ling, R. (**1998**). ''Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies,'' J. Acoust. Soc. Am. **103**, 462–474.

Burns, E. M., Strickland, E. A., Tubis, A., and Jones, K. (**1984**). ''Interactions among spontaneous otoacoustic emissions. I. Distortion products and linked emissions,'' Hear. Res. **16**, 271–278.

Camalet, S., Duke, T., Jülicher, F., and Prost, J. (**2000**). ''Auditory sensitivity provided by self-tuned critical oscillations of hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 3183–3188.

Cleveland, W. S. (**1993**). *Visualizing Data* (Hobart, Summit, NJ).

Cohen, L. (**1995**). *Time-Frequency Analysis* (Prentice Hall PTR, Upper Saddle River, NJ).

Crawford, A. C., and Fettiplace, R. (**1985**). ''The mechanical properties of ciliary bundles of turtle cochlear hair cells,'' J. Physiol. (London) **364**, 359–379.

Dallmayr, C. (**1985**). ''Spontane oto-akustiche Emissionen, Statistik und Reaktion auf akustiche Störtöne,'' Acustica **59**, 67–75.

Davis, H. (**1983**). ''An active process in cochlear mechanics,'' Hear. Res. **9**, 79–90.

Denk, W., and Webb, W. W. (**1992**). ''Forward and reverse transduction at the limit of sensitivity studied by correlating electrical and mechanical fluctuations in frog saccular hair cells,'' Hear. Res. **60**, 89–102.

de Kleine, E., Wit, H. P., van Dijk, P., and Avan, P. (**2000**). ''The behavior of spontaneous otoacoustic emissions during and after postural changes,'' J. Acoust. Soc. Am. **107**, 3308–3316.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). ''Multiple internal reflections in the cochlea and their effect on DPOAE fine structure,'' J. Acoust. Soc. Am. **112**, 2882–2897.

Duke, T. (**2002**). ''The power of hearing,'' Phys. World **15**, 29–33.

Egolf, D. P., Kennedy, W. A., and Larson, V. D. (**1992**). ''Occluded-ear simulator with variable acoustic properties,'' J. Acoust. Soc. Am. **91**, 2813–2823.
Eguíluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A. J., and Magnasco, M. O. (**2000**). ''Essential nonlinearities in hearing,'' Phys. Rev. Lett. **84**, 5232–5235.
Engström, H., Ades, H. W., and Andersson, A. (**1966**). *Structural Pattern of the Organ of Corti* (Williams and Wilkins, Baltimore).
Geisler, C. D. (**1998**). *From Sound to Synapse* (Oxford University Press, New York).
Gleich, O., and Narins, P. M. (**1988**). ''The phase response of primary auditory afferents in a songbird (*Sturnus vulgaris L.*),'' Hear. Res. **32**, 81–91.
Golay, M. J. E. (**1961**). ''Note on coherence vs narrow-bandedness in regenerative oscillators, masers, lasers, etc.,'' Proc. IRE **49**, 958–959.
Gold, T. (**1948**). ''Hearing. II. The physical basis of the action of the cochlea,'' Proc. R. Soc. London, Ser. B **135**, 492–498.
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Hauser, R., Probst, R., and Harris, F. P. (**1993**). ''Effects of atmospheric pressure variation on spontaneous, transiently evoked, and distortion product otoacoustic emissions in normal human ears,'' Hear. Res. **69**, 133–145.
Hillery, C. M., and Narins, P. M. (**1984**). ''Neurophysiological evidence for a traveling wave in the amphibian inner ear,'' Science (Washington, DC, U.S.) **225**, 1037–1039.
Hudspeth, A. J. (**1997**). ''Mechanical amplification of stimuli by hair cells,'' Curr. Opin. Neurobiol. **7**, 480–486.
Jülicher, F., Camalet, S., Prost, J., and Duke, T. A. J. (**2003**). ''Active amplification by critical oscillations,'' in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 16–27.
Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation,'' J. Acoust. Soc. Am. **109**, 622–637.
Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflectance coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.
Kemp, D. T., and Chum, R. A. (**1980**). ''Observations on the generator mechanism of stimulus frequency acoustic emissions—Two tone suppression,'' in *Psychophysical Physiological and Behavioural Studies in Hearing*, edited by G. V. D. Brink and F. A. Bilsen (Delft University Press, Delft), pp. 34–42.
Kemp, D. T. (**1979a**). ''Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea,'' Arch. Oto-Rhino-Laryngol. **224**, 37–45.
Kemp, D. T. (**1979b**). ''The evoked cochlear mechanical response and the auditory microstructure—Evidence for a new element in cochlear mechanics,'' Scand. Audiol. Suppl. **9**, 35–47.
Kemp, D. T. (**1981**). ''Physiologically active cochlear micromechanics—one source of tinnitus,'' in *Tinnitus*, edited by D. Evered and G. Lawrenson (Pitman, London), pp. 54–81.
Kemp, D. T. (**2002**). ''Exploring cochlear status with otoacoustic emissions: The potential for new clinical applications,'' in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 1–47.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). ''Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears,'' J. Acoust. Soc. Am. **109**, 2862–2869.
Köppl, C. (**1995**). ''Otoacoustic emissions as an indicator for active cochlear mechanics: A primitive property of vertebrate auditory organs,'' in *Advances in Hearing Research*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore), pp. 207–218.
Kössl, M. and Vater, M. (**1995**). ''Cochlear structure and function in bats,'' in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 191–234.
Liberman, M. C., Gao, J., He, D. Z. Z., Wu, X., Jia, S., and Zuo, J. (**2002**). ''Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier,'' Nature (London) **419**, 300–304.
Long, G. R., Schaffer, L. A., Dhar, S., and Talmadge, C. L. (**2000**). ''Cross species comparison of otoacoustic fine-structure,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 367–373.
Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1988**). ''Spontaneous otoacoustic emissions in the nonhuman primate. II. Cochlear anatomy,'' Hear. Res. **33**, 69–94.
Lonsbury-Martin, B. L., and Martin, G. K. (**2001**). ''Otoacoustic emissions,'' in *Physiology of the Ear*, edited by A. F. Jahn and J. Santos-Sacchi (Singular, San Diego), pp. 443–480.
Manley, G. A., Yates, G. K., and Köppl, C. (**1990**). ''Peripheral auditory processing in the bobtail lizard Tiliqua rugosa: IV. Phase locking of auditory-nerve fibers,'' J. Comp. Physiol., A **167**, 129–138.
Manley, G. A. (**1990**). *Peripheral Hearing Mechanisms in Reptiles and Birds* (Springer, Berlin).
Manley, G. A. (**2001**). ''Evidence for an active process and a cochlear amplifier in nonmammals,'' J. Neurophysiol. **86**, 541–549.
Martin, P., Hudspeth, A. J., and Jülicher, F. (**2001**). ''Comparison of a hair bundle's spontaneous oscillations with its response to mechanical stimulation reveals the underlying active process,'' Proc. Natl. Acad. Sci. U.S.A. **98**, 14380–14385.
Martin, P., and Hudspeth, A. J. (**1999**). ''Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 14306–14311.
Martin, P., and Hudspeth, A. J. (**2001**). ''Compressive nonlinearity in the hair bundles active response to mechanical stimulation,'' Proc. Natl. Acad. Sci. U.S.A. **98**, 14386–14391.
Martin, P., Jülicher, F., and Hudspeth, A. J. (**2003**). ''The contribution of transduction channels and adaptation motors to the hair cell's active process,'' in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 3–15.
Murphy, W. J., Talmadge, C. L., Tubis, A., and Long, G. R. (**1995a**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. I. Response to pulsed single-tone suppressors,'' J. Acoust. Soc. Am. **97**, 3702–3710.
Murphy, W. J., Tubis, A., Talmadge, C. L., and Long, G. R. (**1995b**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. II. Suppression of interacting emissions,'' J. Acoust. Soc. Am. **97**, 3711–3720.
Murphy, W. J., Tubis, A., Talmadge, C. L., Long, G. R., and Krieg, E. F. (**1996**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. III. Response to a single tone at multiple suppression levels,'' J. Acoust. Soc. Am. **100**, 3979–3982.
Nedzelnitsky, V. (**1980**). ''Sound pressures in the basal turn of the cat cochlea,'' J. Acoust. Soc. Am. **68**, 1676–1689.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**, 652–656.
Neely, S. T. (**1983**). ''The cochlear amplifier,'' in *Mechanics of Hearing*, edited by E. Boer and M. A. Viergever (Martinus Nijhoff, The Hague), pp. 111–118.
Norton, S. J., and Neely, S. T. (**1987**). ''Tone-burst-evoked otoacoustic emissions from normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1860–1872.
Ohyama, K., Wada, H., Kobayashi, T., and Takasaka, T. (**1991**). ''Spontaneous otoacoustic emissions in the guinea pig,'' Hear. Res. **56**, 111–121.
Oxenham, A. J., and Shera, C. A. (**2003**). ''Estimates of human cochlear tuning at low levels using forward and simultaneous masking,'' J. Assoc. Res. Otolaryngol. (in press).
Pang, X. D., and Peake, W. T. (**1986**). ''How do contractions of the stapedius muscle alter the acoustic properties of the ear?,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, Berlin), pp. 36–43.
Peake, W. T., and Ling, A., (**1980**). ''Basilar-membrane motion in the alligator lizard: Its relation to tonotopic organization and frequency selectivity,'' J. Acoust. Soc. Am. **67**, 1736–1745.
Peake, W. T., Rosowski, J. J., and Lynch, T. J. (**1992**). ''Middle-ear transmission: Acoustic versus ossicular coupling in cat and human,'' Hear. Res. **57**, 245–268.
Puria, S., Peake, W. T., and Rosowski, J. J. (**1997**). ''Sound-pressure measurements in the cochlear vestibule of human-cadaver ears,'' J. Acoust. Soc. Am. **101**, 2754–2770.
Puria, S. (**2003**). ''Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions,'' J. Acoust. Soc. Am. **113**, 2773–2789.
Rice, S. O. (**1954**). ''Mathematical analysis of random noise,'' in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York), pp. 133–294.
Russell, A. F. (**1992**). ''Heritability of spontaneous otoacoustic emissions,'' Ph.D. thesis, University of Illinois.

Russell, I. J., and Kössl, M. (**1999**). ''Micromechanical responses to tones in the auditory fovea of the greater mustached bat's cochlea,'' J. Neurophysiol. **82**, 676–686.
Schloth, E., and Zwicker, E. (**1983**). ''Mechanical and acoustical influences on spontaneous otoacoustic emissions,'' Hear. Res. **11**, 285–293.
Schloth, E. (**1983**). ''Relation between spectral composition of spontaneous otoacoustic emissions and fine-structure of threshold in quiet,'' Acustica **53**, 250–256.
Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–2232.
Shera, C. A., and Guinan, J. J. (**2003**). ''Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning,'' J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A., and Zweig, G. (**1991a**). ''Reflection of retrograde waves within the cochlea and at the stapes,'' J. Acoust. Soc. Am. **89**, 1290–1305.
Shera, C. A., and Zweig, G. (**1991b**). ''A symmetry suppresses the cochlear catastrophe,'' J. Acoust. Soc. Am. **89**, 1276–1289.
Shera, C. A., and Zweig, G. (**1992a**). ''Analyzing reverse middle-ear transmission: Noninvasive Gedankenexperiments,'' J. Acoust. Soc. Am. **92**, 1371–1381.
Shera, C. A., and Zweig, G. (**1992b**). ''Middle-ear phenomenology: The view from the three windows,'' J. Acoust. Soc. Am. **92**, 1356–1370.
Shera, C. A., and Zweig, G. (**1993a**). ''Noninvasive measurement of the cochlear traveling-wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.
Shera, C. A., and Zweig, G. (**1993b**). ''Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emission,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.
Siegel, J. H., and Borneman, A. L. (**1999**). ''Comparable measures of perceptual combination tones and distortion-product otoacoustic emissions,'' Assoc. Res. Otolaryngol. Abs. **22**, 393.
Siegman, A. E. (**1986**). *Lasers* (University Science, Sausalito, CA).
Sisto, R., and Moleti, A. (**1999**). ''Modeling otoacoustic emissions by active nonlinear oscillators,'' J. Acoust. Soc. Am. **106**, 1893–1906.
Smolders, J. W. T., and Klinke, R. (**1986**). ''Synchronized responses of primary auditory fiber populations in *Caiman crocodilus* (*L*.) to single tones and clicks,'' Hear. Res. **24**, 89–103.
Smoorenburg, G. F. (**1974**). ''On the mechanisms of combination tone generation and lateral inhibition in hearing,'' in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin), pp. 332–342.
Talmadge, C. L., Long, G. R., Murphy, W. J., and Tubis, A. (**1993**). ''New off-line method for detecting spontaneous otoacoustic emission in human subjects,'' Hear. Res. **71**, 170–182.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Wit, H. P., and Long, G. R. (**1991**). ''Are spontaneous otoacoustic emissions generated by self-sustained cochlear oscillators?,'' J. Acoust. Soc. Am. **89**, 2391–2399.
Talmadge, C. L., and Tubis, A. (**1993**). ''On modeling the connection between spontaneous and evoked otoacoustic emissions,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 25–32.
Tubis, A., and Talmadge, C. L. (**1998**). ''Ear canal reflectance in the presence of spontaneous otoacoustic emissions. I. Limit-cycle oscillator model,'' J. Acoust. Soc. Am. **103**, 454–461.
Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.
van Dijk, P., Manley, G. A., Gallo, L., Pavusa, A., and Taschenberger, G. (**1996**). ''Statistical properties of spontaneous otoacoustic emissions in one bird and three lizard species,'' J. Acoust. Soc. Am. **100**, 2220–2227.
van Dijk, P., Manley, G. A., and Gallo, L. (**1998**). ''Correlated amplitude fluctuations of spontaneous otoacoustic emissions in six lizard species,'' J. Acoust. Soc. Am. **104**, 1559–1564.
van Dijk, P., and Wit, H. P. (**1990a**). ''Amplitude and frequency fluctuations of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **88**, 1779–1793.
van Dijk, P., and Wit, H. P. (**1990b**). ''Synchronization of spontaneous otoacoustic emissions to a $2f_1-f_2$ distortion product,'' J. Acoust. Soc. Am. **88**, 850–856.
van Hengel, P. W. J., Duifhuis, H., and van den Raadt, M. P. M. G. (**1996**). ''Spatial periodicity in the cochlea: The result of interaction of spontaneous emissions?,'' J. Acoust. Soc. Am. **99**, 3566–3571.
Wilson, J. P., and Sutton, G. J. (**1981**). ''Acoustic correlates of tonal tinnitus,'' in *Tinnitus*, edited by D. Evered and G. Lawrenson (Pitman, London), pp. 82–107.
Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure, and tonal tinnitus,'' Hear. Res. **2**, 233–252.
Wright, A. A. (**1984**). ''Dimensions of the cochlear stereocilia in man and in guinea pig,'' Hear. Res. **13**, 89–98.
Zurek, P. M. (**1981**). ''Spontaneous narrow-band acoustic signals emitted by human ears,'' J. Acoust. Soc. Am. **69**, 514–523.
Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.
Zweig, G. (**1991**). ''Finding the impedance of the organ of Corti,'' J. Acoust. Soc. Am. **89**, 1229–1254.
Zweig, G. (**2003**). ''Cellular cooperation in cochlear mechanics,'' in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 315–330.
Zwicker, E., and Harris, F. P. (**1990**). ''Psychoacoustical and ear canal cancellation of $(2f_1-f_2)$-distortion products,'' J. Acoust. Soc. Am. **87**, 2583–2591.
Zwicker, E., and Manley, G. (**1981**). ''Acoustical responses and suppression-period patterns in guinea pigs,'' Hear. Res. **4**, 43–52.
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 otoacoustic emissions,'' J. Acoust. Soc. Am. **75**, 1148–1154.
Zwicker, E. (**1988**). ''The inner ear, a sound processing and a sound emitting system,'' J. Acoust. Soc. Jpn. **9**, 59–74.
Zwicker, E. (**1989**). ''Otoacoustic emissions and cochlear travelling waves,'' in *Cochlear Mechanisms—Structure, Function, and Models*, edited by J. P. Wilson and D. T. Kemp (Plenum, New York), pp. 359–366.

Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears

Michael P. Gorga,[a] Stephen T. Neely, Darcia M. Dierking, Patricia A. Dorn, Brenda M. Hoover, and Denis F. Fitzpatrick
Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 17 November 2002; revised 20 March 2003; accepted 21 March 2003)

Distortion product otoacoustic emission (DPOAE) suppression measurements were made in 20 subjects with normal hearing and 21 subjects with mild-to-moderate hearing loss. The probe consisted of two primary tones (f_2, f_1), with f_2 held constant at 4 kHz and $f_2/f_1 = 1.22$. Primary levels (L_1, L_2) were set according to the equation $L_1 = 0.4L_2 + 39$ dB [Kummer *et al.*, J. Acoust. Soc. Am. **103**, 3431–3444 (1998)], with L_2 ranging from 20 to 70 dB SPL (normal-hearing subjects) and 50–70 dB SPL (subjects with hearing loss). Responses elicited by the probe were suppressed by a third tone (f_3), varying in frequency from 1 octave below to $\frac{1}{2}$ octave above f_2. Suppressor level (L_3) varied from 5 to 85 dB SPL. Responses in the presence of the suppressor were subtracted from the unsuppressed condition in order to convert the data into decrements (amount of suppression). The slopes of the decrement versus L_3 functions were less steep for lower frequency suppressors and more steep for higher frequency suppressors in impaired ears. Suppression tuning curves, constructed by selecting the L_3 that resulted in 3 dB of suppression as a function of f_3, resulted in tuning curves that were similar in appearance for normal and impaired ears. Although variable, Q_{10} and Q_{ERB} were slightly larger in impaired ears regardless of whether the comparisons were made at equivalent SPL or equivalent sensation levels (SL). Larger tip-to-tail differences were observed in ears with normal hearing when compared at either the same SPL or the same SL, with a much larger effect at similar SL. These results are consistent with the view that subjects with normal hearing and mild-to-moderate hearing loss have similar tuning around a frequency for which the hearing loss exists, but reduced cochlear-amplifier gain.
[DOI: 10.1121/1.1575751]

PACS numbers: 43.64.Ha, 43.64.Jb [BLM]

I. INTRODUCTION

Distortion-product otoacoustic emissions (DPOAE) are a byproduct of normal nonlinear cochlear processes. Their generation is linked to the status of the outer hair cells (OHC). Recent findings, in which DPOAE data were collected in mice for which the putative ''motor'' molecule (prestin) was knocked out, are consistent with the view that DPOAEs are generated by OHC motility (Liberman *et al.*, 2002). Under appropriate stimulus conditions, the normal-hearing ear (having normal OHC function) produces these distortion products in response to pairs of primary tones. Because OHC damage results in a reduction or loss of normal nonlinear behavior (e.g., Dallos *et al.*, 1980; Kim, 1980), one manifestation of damage to the normal nonlinear system (i.e., the OHC system) is the reduction or loss of DPOAEs. It also is known that damage to OHCs results in auditory threshold elevation.

The relationships among OHC status, auditory function, and DPOAE levels have led to the application of DPOAE measurements to the task of identifying hearing loss. Several studies have described DPOAE measurements in normal-hearing and hearing-impaired ears (e.g., Martin *et al.*, 1990; Gorga *et al.*, 1993, 1996, 1997, 2000; Kim *et al.*, 1996). These studies have shown that DPOAEs accurately identify auditory status at mid- and high frequencies, performing less accurately at lower frequencies. These efforts have focused primarily on the ability of DPOAEs to identify auditory status, as defined by the pure-tone audiogram. As such, they have focused on dichotomous decisions in which DPOAE measurements were used to classify an ear as having either normal hearing or hearing impairment. To a lesser extent, others have described the relation between behavioral thresholds and DPOAEs, by relating DPOAE threshold and auditory thresholds (Martin *et al.*, 1990; Gorga *et al.*, 1996; Boege and Janssen, 2002; Gorga *et al.*, 2003), or by relating DPOAE level and auditory threshold (Martin *et al.*, 1990; Dorn *et al.*, 2001; Gorga *et al.*, 2002a).

Threshold elevation, however, is only one of several consequences of OHC damage. There are both mechanical and neural data from lower animals, indicating that frequency selectivity may be reduced and response growth may become more rapid when OHC damage exists (e.g., Evans, 1974; Kiang *et al.*, 1976; Dallos and Harris, 1978; Liberman and Dodds, 1984; Sewell, 1984; Gorga and Abbas, 1981b; Ruggero and Rich, 1991). Frequency selectivity in these studies typically is defined by the parametrization of tuning curves, using measures such as Q_{10} and differences in threshold at the tip and on the tail of tuning curves. Response growth refers to the rate at which either discharge rate (for single-unit studies), displacement or velocity (for basilar-membrane studies), or masking (for whole-nerve action po-

[a]Electronic mail: gorga@boystown.org

tential studies) grows as stimulus level is increased. These ''suprathreshold'' effects may be related to perceptual phenomena such as the reduced speech-perception abilities or loudness recruitment that sometimes are associated with peripheral hearing loss. For example, compensating for abnormally rapid growth of loudness is the motivation for including compression in many hearing aids. In one sense, the compression circuit of the hearing aid is attempting to compensate for the loss of the compressive behavior of the normal cochlea.

While direct measurements of frequency selectivity or response growth are not possible in humans, indirect measures may provide insight into underlying properties related to cochlear processing. Psychoacoustic and electrophysiologic masking studies are motivated by a desire to indirectly estimate representations of cochlear properties that have been identified from more direct, invasive measurements in lower animals. Recently, for example, Oxenham and Plack (1997) related the growth of forward masking (measured behaviorally) to nonlinear cochlear processing, based on a comparison of on-frequency and low-frequency masking effects.

DPOAE measurements in a suppression paradigm have also been used to describe peripheral response properties in humans. Several studies have shown that DPOAE suppression measurements provide information related to the site of generation for DPOAEs and the tuning properties of the cochlea at the place(s) of DPOAE generation, at least for ears with normal cochleae (e.g., Brown and Kemp, 1984; Martin *et al.*, 1987). Furthermore, studies in humans have shown that the reductions in DPOAE levels in the presence of a suppressor (Abdala, 1998, 2001; Abdala *et al.*, 1996; Kummer *et al.*, 1995; Gorga *et al.*, 2002b) share a dependence on frequency that is similar to other measures of response growth such as single-unit rate-level functions (Sachs and Abbas, 1974; Schmiedt and Zwislocki, 1980) and measurements of basilar-membrane motion (Ruggero and Rich, 1991; Ruggero *et al.*, 1997), that were based on more direct measurements in lower animals. These similarities have led us to pursue DPOAE suppression measurements in patients with cochlear hearing loss in order to determine whether changes in response growth and tuning, evident in direct measurements from lower animals, can be indirectly observed in humans using the same noninvasive DPOAE techniques that have been applied with humans having normal auditory thresholds.

The present study describes our initial efforts to measure DPOAE suppression in human subjects with mild-to-moderate hearing loss. Subjects with normal hearing are included for comparison purposes. Of primary interest was determining whether reduced frequency selectivity and/or more rapid response growth, predicted from single-unit and basilar-membrane studies in lower animals, can be observed in indirect DPOAE measurements in humans. A long-term goal of our research program has always been to determine whether there are relationships between objective measurements of cochlear function (such as DPOAE and/or evoked potential (EP) measurements) and perceptual consequences of damage to the cochlea (such as threshold elevation or abnormal growth of loudness) so that eventually the perceptual consequences can be predicted from the objective measures. This long-term goal has clinical importance related to the identification and the rehabilitation of hearing loss (such as the selection of hearing-aid characteristics) in infants and young children who may be unable to provide behavioral responses to sound, especially subjective judgments such as loudness. While a great deal is already known about the relationship between DPOAE measurements and sensitivity loss (e.g., Martin *et al.*, 1990; Gorga *et al.*, 1993, 1997, 2000; Stover *et al.*, 1996; Kim *et al.*, 1996; Boege and Janssen, 2002), the present work is viewed as an initial step towards understanding the relationship between DPOAE measurements and suprathreshold consequences of cochlear damage. The primary aim is to determine whether DPOAE measurements of frequency selectivity and response growth are possible in subjects having hearing loss. This initial step is needed prior to efforts to correlate DPOAE data with suprathreshold perceptual phenomena because of the tendency for DPOAEs to be reduced or absent when hearing loss exists. If DPOAE suppression studies can be performed in humans with hearing loss, future work will determine the extent to which these measurements correlate with behavioral findings.

II. METHODS

A. Subjects

Twenty-two ears of 20 subjects with normal hearing and 29 ears of 25 subjects with hearing loss participated in these studies. The two subject groups differed in age, with the normal-hearing subjects being younger (mean=24.9 years, s.d.=13.4 years) than the group of subjects with hearing loss (mean=57.3 years, s.d.=14.6 years). It was difficult to recruit older normal-hearing subjects or younger hearing-impaired subjects, due to factors associated with subject availability. Thus, there is concern that some of the results reviewed below might have been influenced by age rather than hearing loss. As will be seen, however, it is not obvious how age could have affected the data in such a way as to result in the outcomes that were observed.

Normal hearing was defined as audiometric thresholds less than or equal to 15 dB HL (*re*: ANSI, 1996) at 4 kHz. The mean threshold for subjects in the normal-hearing group at 4 kHz (the probe frequency; see below) was 5.7 dB HL (s.d.=4.4 dB). Subjects with thresholds of 20 dB HL or greater at 4 kHz were classified as hearing impaired. It was assumed that the hearing losses in these subjects were of cochlear origin, based upon their case histories, pure-tone audiometry, acoustic immittance measures, and prior results of any additional diagnostic tests. However, a precise diagnosis of etiology was not available for the majority of the subjects with hearing loss. This ''diagnostic ambiguity'' highlights one of the issues that is confronted whenever studies are conducted in humans. Whereas in animal studies, cochlear damage is typically induced and, therefore, controlled, such control is frequently impossible in studies that involve humans with hearing loss. In fact, it is difficult to find a homogeneous group of human subjects with hearing loss.

Subjects with hearing loss were selected according to their thresholds at 4 kHz, which was the f_2 frequency that was used as the "probe frequency" during the DPOAE suppression experiments. Special efforts were made to include subjects for whom the magnitude of the loss was between 20 and 55 dB HL, on the assumption that it would be more likely that DPOAE level would be too reduced or the response would be completely absent for ears with hearing losses exceeding the upper limit of this range. Even so, DPOAE suppression experiments were not possible in four of the 25 subjects with hearing loss because they did not produce responses of sufficient level to permit reliable estimates of the suppression of that response. Thus, the data reported below are based on measurements from 23 ears of 21 subjects with mild-to-moderate hearing loss, who represent a subset of the potential subjects with hearing loss in this range. The mean thresholds at 4 kHz for the group of subjects with hearing loss was 36.3 dB HL (s.d.=9.1 dB). Thus, on average, the subjects with hearing loss had thresholds that were 30.6 dB higher than thresholds for the normal-hearing group.

B. Stimuli

DPOAE data were collected with custom-designed software (EMAV, Neely and Liu, 1994). This system controlled a sound card (CardDeluxe, Digital Audio Labs) that was housed in a PC. Separate channels of the sound card were used to output the two primary tones (f_1 and f_2) that were mixed acoustically in the ear canal. The channel that was used to generate f_2 was also used to generate the suppressor tone (f_3). The output of the sound card was delivered to a probe–microphone system (Etymotic, ER 10C). This system was modified in order to remove 20 dB of attenuation from each channel, thus permitting the presentation of suppressors at higher levels than would be otherwise possible. Separate loudspeakers in the probe were used to transduce the outputs from the two separate channels of the sound card. The probe's microphone was used to measure levels at the plane of the probe.

All DPOAE data were collected for the condition in which f_2 was fixed at 4 kHz, $f_2/f_1=1.22$, the level of f_2 (L_2) was varied from 20 to 70 dB SPL (in 10-dB steps), and, for each L_2, L_1 was set according to the equation, $L_1=0.4L_2+39$ dB (Kummer *et al.*, 1998). A third tone, f_3, was used to suppress the response elicited by the primary tones. Seventeen suppressor frequencies were used, ranging from 1 octave below (2 kHz) to $\frac{1}{2}$ octave above (5.6 kHz) f_2. Suppressor level (L_3) was varied in 5-dB steps from 5 dB SPL up to a maximum level of 85 dB SPL.

Prior to data collection, real-ear measurements at the plane of the probe were used to calibrate the stimuli. This calibration procedure may introduce errors, especially for frequencies close to the probe frequency used in the present experiment ($f_2=4$ kHz). These errors are a consequence of standing-wave problems introduced by an interaction between the quarter wavelength of the probe frequency and the dimensions of adult ear canals with a probe system in place (Siegel, 1994, 2002). However, the use of standard cavities for calibration also introduces errors related to the fact that these cavities may not provide a good model for individual ears. The best solution might be one in which acoustic intensity is measured (Neely and Gorga, 1998). That approach, however, has not been implemented in any widely available device or software system. Thus, while it is recognized that real-ear calibration measurements may introduce some errors, it is viewed as an acceptable compromise, given the current state of calibration methods.

C. Procedures

Both DPOAE and noise levels were estimated from the energy in the $2f_1-f_2$ frequency bin. During data collection, each 2-s sample was alternately stored in one of two buffers. In order to estimate DPOAE level, the contents of the two buffers were summed. Noise level was estimated by subtracting the contents of one buffer from the contents of the other buffer. This approach is attractive in that signal and noise are estimated within the same frequency bin. However, it has the disadvantage of providing a more variable estimate of noise level, compared to the case when noise levels are estimated from the contents of several bins adjacent to the $2f_1-f_2$ (signal) frequency bin.

Measurement-based stopping rules were used during DPOAE measurements. For each condition, averaging continued (1) until the noise floor was -25 dB SPL or less, or (2) for 32 s of artifact-free averaging, whichever occurred first. The noise-stopping rule allowed us to measure DPOAEs (and, therefore, suppression) over a wide dynamic range and still be confident that measured responses were above the level at which system distortion occurred (see Dorn *et al.*, 2001 for a more complete description of system distortion for the present measurement system). The time limit prevented data collection from continuing indefinitely for any single stimulus condition.

In each subject, a DPOAE input/output (I/O) function was measured when $f_2=4$ kHz. These initial measurements were needed in order to select the range of L_2 levels for each subject over which the suppression experiments could be conducted. Figure 1 shows mean DPOAE I/O functions for both normal-hearing and hearing-impaired subjects. The mean noise floor is also shown. Note that DPOAE levels were less in impaired ears at the same L_2 levels. The response levels at 50–70 dB SPL in impaired ears were more like the levels observed at 20–40 dB SPL in normal ears. Significant differences in DPOAE level were observed when normal and impaired responses were compared at the same SPL. Differences were not significant when the levels in impaired ears at 50–70 dB SPL were compared to the levels observed in normal ears at 20–40 dB SPL. That is, if the results are shifted in impaired ears by 30 dB (an amount that is nearly equivalent to the mean behavioral threshold difference between groups), the levels produced by both groups were similar. Another way of thinking about these findings is that normal and impaired ears produced different DPOAE levels when comparisons were made at similar SPL, but not when comparisons were made at levels that were approximately equivalent in terms of sensation level (SL). We will return to this SPL/SL comparison when considering the primary findings of the present study.

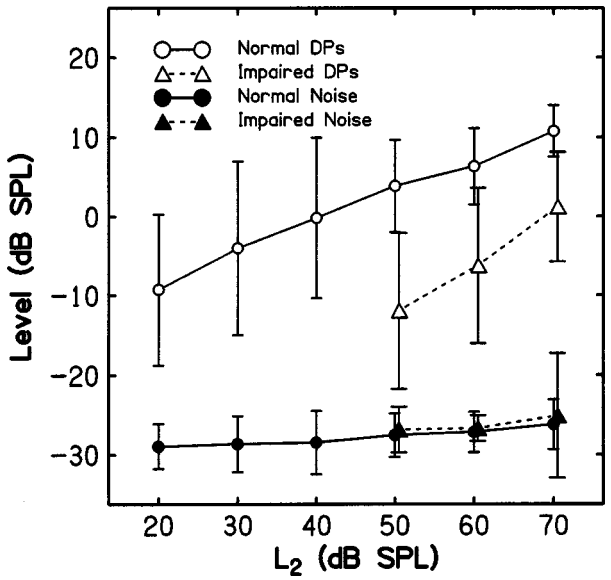


FIG. 1. Mean DPOAE and noise level (dB SPL) as a function of L_2 (dB SPL). Circles represent data from normal ears, while triangles represent data from ears with hearing loss. Open and filled symbols represent DPOAE and noise levels, respectively. Error bars represent 1 s.d. Data points were offset slightly in the L_2 dimension to help visualization.

The influence of the stopping rule is evident in the constancy of noise levels across L_2 and between the two subjects groups. For each subject, suppression experiments were performed at those L_2 levels for which at least a 10-dB signal-to-noise ratio (SNR) was evident in the previously measured I/O functions. As can be seen in Fig. 1, sufficient DPOAE level was measured over the range of L_2 levels from 20 to 70 dB SPL in subjects with normal hearing. The range of levels was more restricted in subjects with hearing loss, being limited to L_2 levels of 50–70 dB SPL.

In each set of suppression measurements, L_2 was fixed at one of six levels in normal ears (20–70 dB SPL, 10-dB steps) or one of three levels in impaired ears (50–70 dB SPL). Experimental conditions were then selected such that f_3 was fixed and its level was varied from 5 to 85 dB SPL (5-dB steps). DPOAE levels were measured for each of these 17 suppressor levels at each f_3 frequency. However, the initial condition prior to the presentation of each f_3 was a control condition, in which no suppressor was presented. Once an L_3 series had been completed for one f_3, a different f_3 was selected and the entire process was repeated. This approach continued until measurements were completed for each of 17 suppressor frequencies. For each subject, no more than 12 hours of data-collection time was required, depending mainly on the range of L_2 levels for which reliable DPOAEs could be measured. This means that data-collection time was greater for subjects with normal hearing (for whom measurements were possible at up to six L_2 levels), but less time was required for subjects with hearing loss.

Following data collection, DPOAE levels were converted to decrements (or amount of suppression) by subtracting the level measured in each suppressor condition from the level measured in the preceding control condition. There were several reasons why the data were converted into decrements. Decrements have the advantage of converting the DPOAE data into amount of suppression (in dB), and they partially account for variance in absolute DPOAE level across subjects. Decrement-versus-L_3 functions for each f_3 were fit with a linear equation in order to provide an estimate of response growth to the suppressor at the place where (presumably) the $2f_1-f_2$ distortion product was initially generated. The fits were restricted to conditions in which the SNR$\geqslant$3 dB, at least one decrement was required to be in the range from 3 to 15 dB, and, outside of this range, decrements were required to increase monotonically with respect to L_3. This combination of inclusion criteria prevented the inclusion of any conditions in the fits in which no suppression occurred or in which the response was completely suppressed. In an effort to include data points in the linear fits for which the SNR was high (i.e., points for which the decrement was as little as 1 dB), these data were transformed by the equation

$$D = 10\log(10^{\mathrm{decr}/10}-1). \tag{1}$$

When $D=0$, the decrement$=$3 dB. This transformation had the effect of linearizing the decrement functions. Finally, in order to reduce the influence of points with low SNRs, data points were weighted by the following equation:

$$\text{SNR weight} = (10^{\mathrm{signal}/10})/(10^{\mathrm{signal}/10}+10^{\mathrm{noise}/10}). \tag{2}$$

It should be noted, however, that there was little difference between weighted and unweighted fits to the data. In fact, the two fits superimposed for the majority of conditions. Figure 2 summarizes the effects of the above treatments to the data for one subject with normal hearing. Circles represent the unprocessed decrements with signal-to-noise ratios (SNR)$\geqslant$3 dB, triangles represent the same data after transformation by the above equation, and the lines represent the weighted fits to the transformed data. The small filled circles represent conditions not meeting the SNR criterion (SNR$\leqslant$3 dB). Thus, they represent conditions in which the DPOAE was at a level that was nearly or completely suppressed. The top, middle, and bottom panels show the data for the cases in which the suppressor (f_3) was below (2.2 kHz), close to (4.1 kHz), and above (4.8 kHz) f_2. As can be seen, the transformation's only effect was to extend the usable range of data, and the weighted functions provide good fits to the data. These trends were evident for decrement functions in both normal-hearing and hearing-impaired ears. The linear equations were solved for the suppressor level that resulted in a decrement of 3 dB (i.e., $D=0$, or 3 dB of suppression). These "threshold" decrements were plotted as a function of f_3 in order to generate DPOAE suppression tuning curves (STC). Because of the differences in L_2 conditions between subject groups, these tuning curves were constructed for L_2 levels from 20–70 dB SPL in normal-hearing subjects, but were restricted to probe levels of 50–70 dB SPL in subjects with hearing loss, due to the smaller responses for the control condition in hearing-impaired ears.

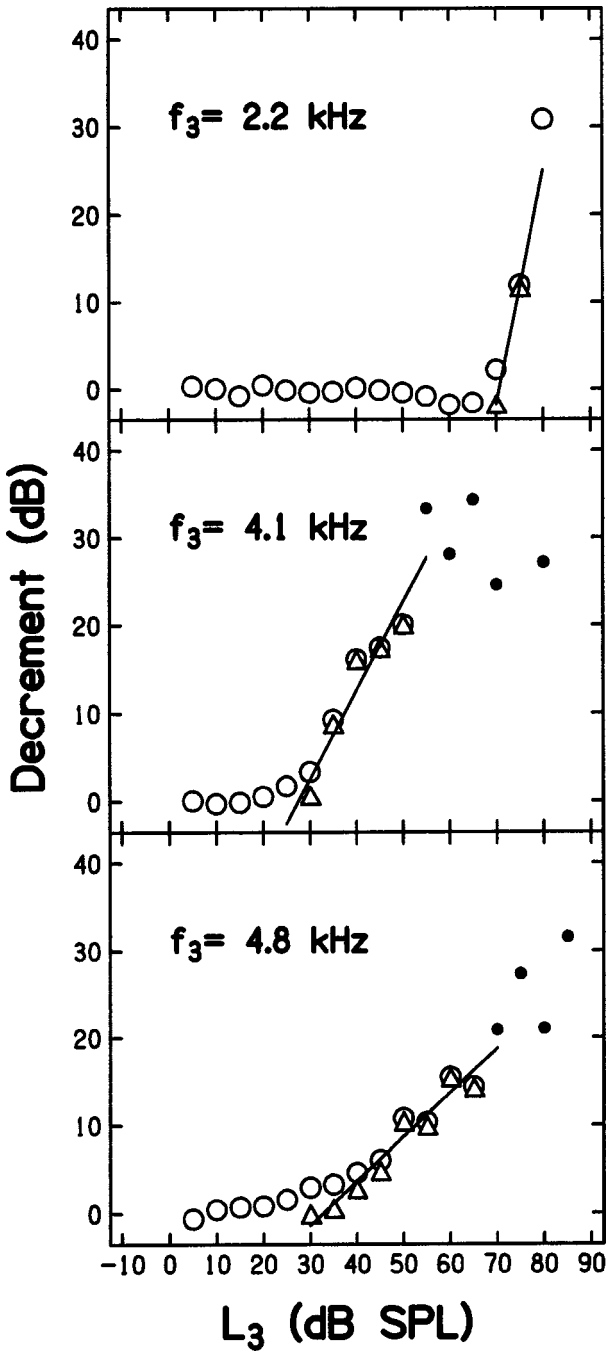


FIG. 2. Decrement (in dB) as a function of suppressor level (L_3) for a low-frequency (2.2 kHz, top), on-frequency (4.1 kHz, middle), and high-frequency suppressor (4.8 kHz, bottom), relative to f_2 for one normal-hearing subject. $L_2=30$ dB SPL. Open circles represent the decrements for which the SNR was at least 3 dB, small filled circles represent data for cases in which the SNR criterion was not met, triangles represent the transformed decrement data meeting the SNR criterion (see the text), and the lines represent linear fits after weighting data according to the SNR. Line fits were restricted according to the rules described in the text.

III. RESULTS

A. Decrement versus suppressor-level functions in normal-hearing ears

Figure 3 shows mean decrement-versus-L_3 functions for subjects with normal hearing. The left column represents data for the case when the suppressor was lower in frequency than the probe ($f_3<f_2$), while the right column represents data for the opposite case ($f_3>f_2$). Each row represents data for a different probe level (L_2). The heavy line represents data for the f_3 frequency closest to f_2 for the set of f_3 frequencies shown in each panel. Thinning lines represent data for f_3 frequencies that are increasingly distant from f_2. Consistent with previous data describing DPOAE suppression (e.g., Martin *et al.*, 1987; Kummer *et al.*, 1995; Abdala *et al.*, 1996; Gorga *et al.*, 2002b), suppression occurred at the lowest levels when f_3 was close to f_2, with increasing suppression "thresholds" as f_3 moved away from f_2 in either direction. In addition, the functions for lower f_3 frequencies relative to f_2 were steeper, compared to the functions when $f_3>f_2$. These trends were evident at all L_2 levels, although it became progressively more difficult to suppress the DPOAE as L_2 increased. An estimate of the dynamic range of the measurements can be obtained by examining the level at which the decrement functions "saturate." When this occurs, the DPOAE was completely suppressed into the noise floor. Examination of these decrement functions indicate that the dynamic range of the present measurements varied from about 20 dB ($L_2=20$ dB SPL) up to more than 30 dB ($L_2=50$, 60, and 70 dB SPL).

B. Decrement versus suppressor-level functions in hearing-impaired ears

In similar fashion, Fig. 4 presents mean decrements as a function of L_3 for subjects with hearing loss. In all respects, the conventions followed in this figure are identical to those used in Fig. 3. However, data collection was restricted to L_2 levels of 50 dB SPL or higher. Decrement functions from impaired ears tended to saturate at lower levels, compared to similar conditions in subjects with normal hearing. Thus, the dynamic range of these measurements was reduced, which probably is due to the same mechanisms that resulted in the reduced L_2 range (of unsuppressed conditions) over which measurable DPOAEs were observed. With the exception of differences in terms of the probe levels (L_2's) for which the experiments could be performed and the reduced dynamic range, there were no apparent differences between these decrement functions and those observed in subjects with normal hearing. That is, thresholds were lowest for f_3 frequencies close to f_2, and the decrement functions appear steeper when $f_3<f_2$, compared to the case when $f_3>f_2$.

C. Slopes of decrement versus suppressor-level functions

Each decrement versus L_3 function was fit with a linear regression in order to provide an estimate of the slope of these functions, as described previously. Figure 5 plots the mean slopes from these linear regressions as a function of f_3. In all cases, solid lines are used to depict the estimated slopes from ears with normal hearing and dotted lines are used to represent the data from ears with hearing loss. Each panel shows data for a different L_2. For the reasons described above, slope could be estimated for L_2 levels ranging from 20 to 70 dB SPL in subjects with normal hearing, but could be estimated only for levels from 50 to 70 dB SPL in subjects with hearing loss.

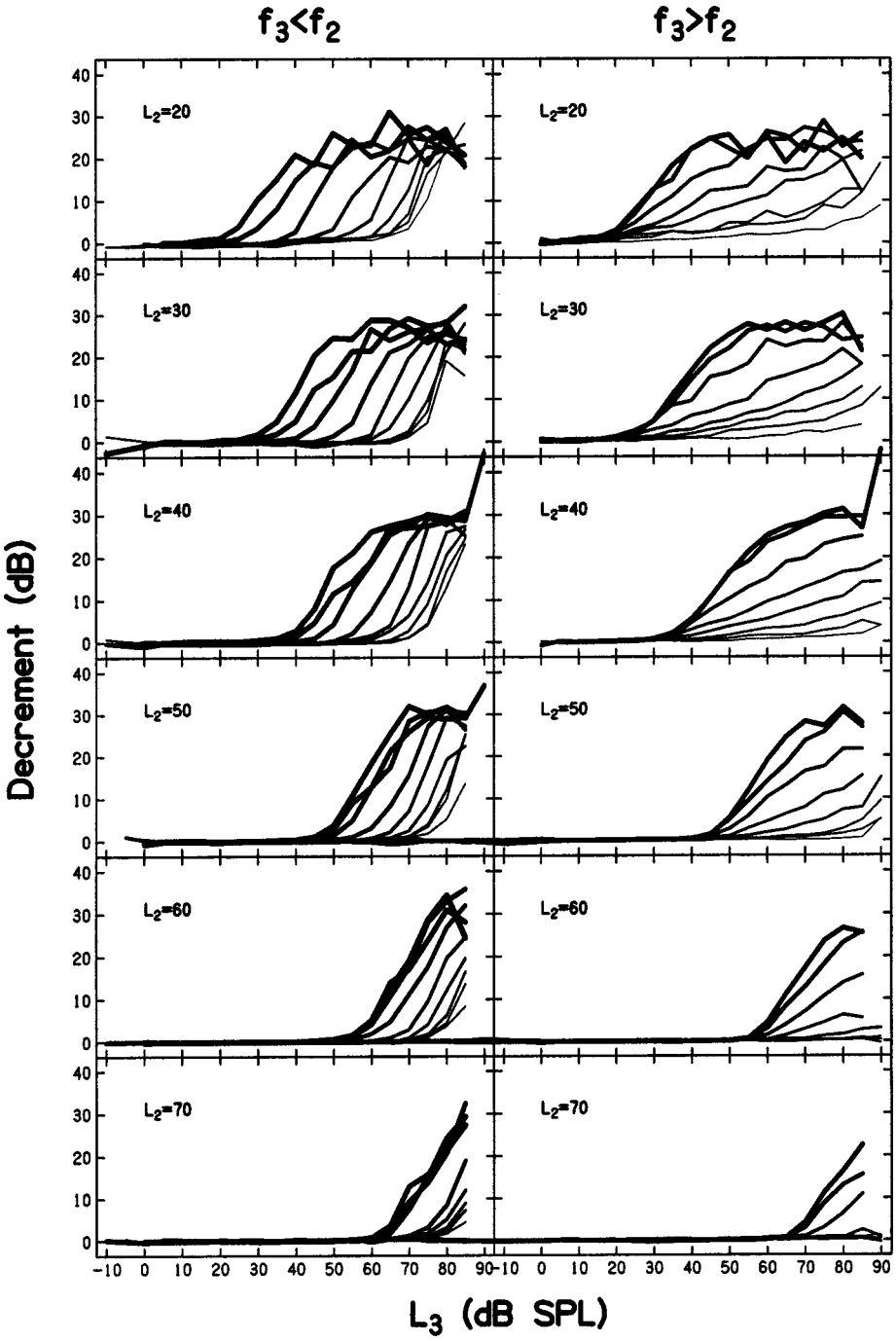


FIG. 3. Mean decrements (in dB) as a function of suppressor level (L_3) for subjects with normal hearing. The left column shows data for the case when the suppressor frequency (f_3) was less than the probe frequency (f_2), while the right column represents the results when f_3 was greater than f_2. Each row presents data for a different probe level (L_2), ranging from 20 to 70 dB SPL. Within each panel, the heavy lines represent decrements for the f_3 closest to f_2 of the set of f_3 frequencies represented in each panel. Thinning lines represent data for f_3 frequencies increasingly more distant from f_2.

These slope estimates were variable across f_3, but tended to decrease as f_3 increased both for normal and impaired ears. For subjects with normal hearing, the slopes were generally greater than 1 for f_3 frequencies below f_2, decreasing rapidly as f_3 moved towards frequencies just above f_2. A similar pattern was observed in subjects with hearing loss, although the apparent differences in slope as a function of frequency were less. Shallower slopes were observed for the highest f_3 frequencies in both subject groups; however, the functions were not always monotonic with frequency. The slopes of decrement functions in subjects with hearing loss appeared to be slightly shallower on the low-frequency side and slightly steeper on the high-frequency side, compared to similar data from ears with normal hearing.

In order to analyze these frequency effects on slope of the decrement functions, frequency was divided into two groups, $f_3<f_2$ and $f_3>f_2$. Within these groups, the slopes for individual f_3 frequencies were averaged to provide single estimates for low- and high-frequency suppressors. In addition, stimulus level was treated in two different ways. In the first analysis, average slopes for normal and impaired ears were compared when stimulus level (L_2) was constant for the two groups. This means that data were compared when L_2 was 50–70 dB SPL for both groups of subjects. In a second analysis, data for normal ears at L_2 levels of 20–40 dB SPL were compared to data from impaired ears when $L_2=50-70$ dB SPL. Two observations from the present study provide support for applying this 30-dB shift in the range of levels over which comparisons between the two

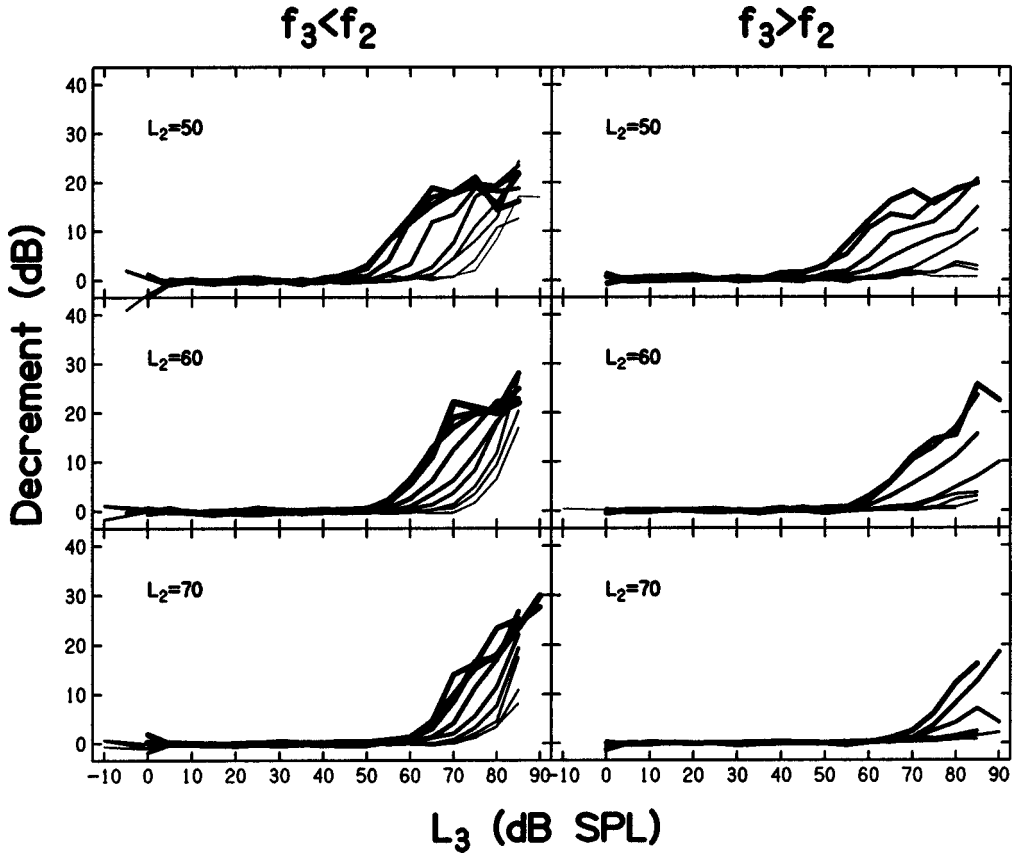


FIG. 4. Same as Fig. 3, only here data are shown for subjects with hearing loss. Note that data are shown only for L_2 levels of 50, 60, and 70 dB SPL, for the reasons described in the text.

groups were made. Recall that the mean normal audiometric threshold was 5.7 dB HL while the mean thresholds for the subjects with hearing loss was 36.3 dB HL, a 30.6-dB difference. In addition, the DPOAE levels differed between the two groups when these levels were compared at the same SPL, but did not differ when the levels produced by impaired ears at 50–70 dB SPL were compared to the levels produced by the normal-hearing group at 20–40 dB SPL, a 30-dB shift for the normal ears (see Fig. 1). Thus, we consider the 30-dB shift as something akin to making the comparison at equivalent SL, at least on average.

Analysis of variance (ANOVA) was used to evaluate the effects of frequency (either above or below f_2), level, and hearing-status group on estimates of slope. For both equivalent SPL and equivalent SL, the outcomes of the ANOVA were the same. Significant effects were observed for frequency and for a frequency×hearing-status group interaction. That is, the slopes were steeper for suppressors below f_2 compared to the slopes for suppressors above f_2 in both groups, but the differences depended on hearing status. Subjects with normal hearing had greater differences between slopes above and below f_2, compared to subjects with hearing loss. This occurred because there was less difference in slope for low- and high-frequency suppressors in subjects with hearing loss. This observation indicates that, at the place where the DPOAE was generated, the growth of response for off-frequency and on-frequency stimuli was more similar in ears with hearing loss.

D. DPOAE suppression tuning curves (STC)

For each L_2, the linear regressions were solved for the L_3 that resulted in 3 dB of suppression for each f_3 frequency. These levels were then plotted as a function of f_3 to produce DPOAE STCs, which are shown in Fig. 6 for one subject with normal hearing and three subjects with hearing loss. The parameter in each panel is L_2. The symbols represent the L_3 necessary for 3 dB of suppression for each f_3. Also shown are the values of Q_{10}, Q_{ERB}, and tip-to-tail differences for each STC. *Q*s were estimated from spline fits to the data, while tip-to-tail differences represent the dB differences between suppression thresholds at $f_3=2.2$ kHz and $f_3=4.1$ kHz. In the normal-hearing subject, sharp tuning was evident around the tip, and the difference between thresholds at the tip and on the low-frequency tail was 40 dB when $L_2=20$ dB SPL. For this subject, there was a systematic decrease in Q_{10}, Q_{ERB}, and tip-to-tail differences as L_2 increased, primarily as a result of changes in suppression threshold around the tip. Less change was evident on the tail. Data from three subjects with hearing loss are shown in the next three panels. While the data from impaired ears were less orderly compared to the data from the subject with normal hearing, the general patterns were similar. At the same SPL, the STCs from the impaired ears appeared similar, and *Q*- and tip-to-tail values were grossly in the range observed for the subject with normal hearing.

Figure 7 shows mean STCs for both normal-hearing and

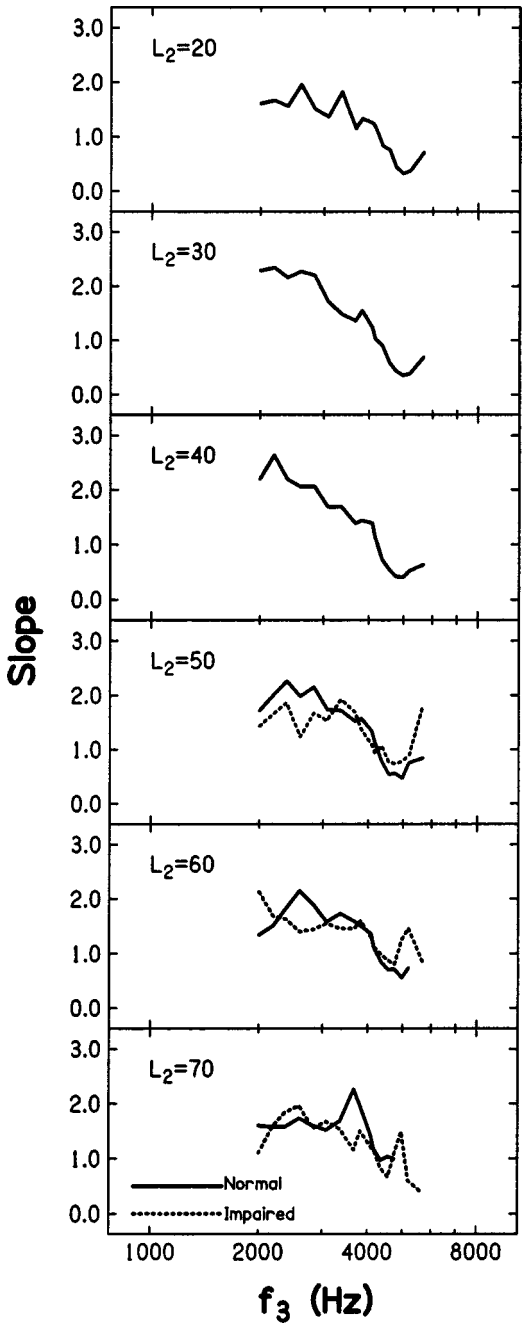


FIG. 5. Mean slopes of decrement versus suppressor level (L_3) functions as a function of suppressor frequency (f_3). Each panel represents data for a different L_2. Within each panel, data from normal-hearing subjects are shown as a solid line, and data from subjects with hearing loss are shown as a dotted line.

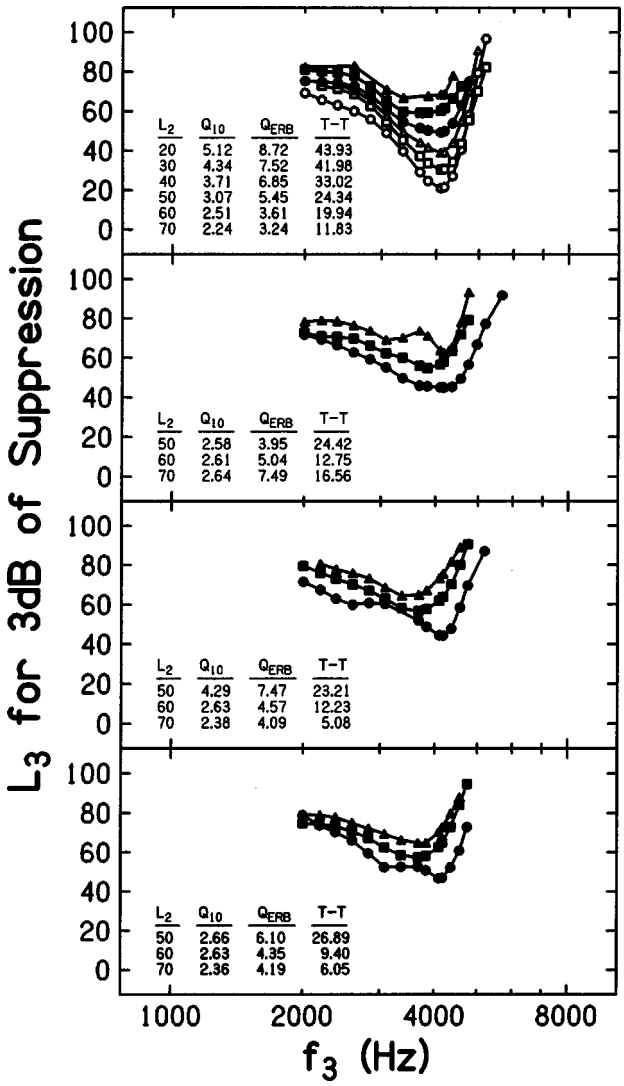


FIG. 6. Suppressor level (L_3) necessary for 3 dB of suppression as a function of f_3 for a subject with normal hearing (top) and for three subjects with hearing loss (bottom three panels). Within each panel, the parameter is L_2. Q_{10}, Q_{ERB}, and tip-to-tail differences for each tuning curve are given within each panel.

hearing-impaired subjects. Each panel represents data for a different L_2, with solid lines representing data for normal ears, and dotted lines representing data for ears with hearing loss. As expected from the previous discussion, DPOAE STCs could be constructed for L_2 levels ranging from 20 to 70 dB SPL in ears with normal hearing and from 50 to 70 dB SPL in ears with hearing loss.

STCs in normal and impaired ears appeared to be similar at the three absolute levels for which comparisons could be made. In the STCs from ears with normal hearing, there was a tendency for the best suppressor frequency to shift towards lower frequencies as level increased. Although less obvious in ears with hearing loss (primarily because there was a smaller range of L_2 levels to consider), the best frequency also shifted towards lower frequencies as L_2 increased. There was a tendency for the suppression thresholds on the tails of the STCs to occur at a lower level for ears with hearing loss, compared to thresholds for similar frequencies in normal ears.

E. Q_{10} and Q_{ERB}

Figure 8 provides a scatter plot of the individual values of Q_{10} and Q_{ERB} (top and bottom rows, respectively) as a function of audiometric threshold. Q_{10} is defined as the best frequency (f_3 frequency with the lowest suppression threshold) divided by the bandwidth at a level 10 dB above the suppression threshold at best frequency (BF). Q_{ERB} is defined as the BF divided by the equivalent rectangular bandwidth (ERB). For any filter, the corresponding ERB is the bandwidth of the rectangular filter with the same BF response that passes the same total power. The DPOAE STCs

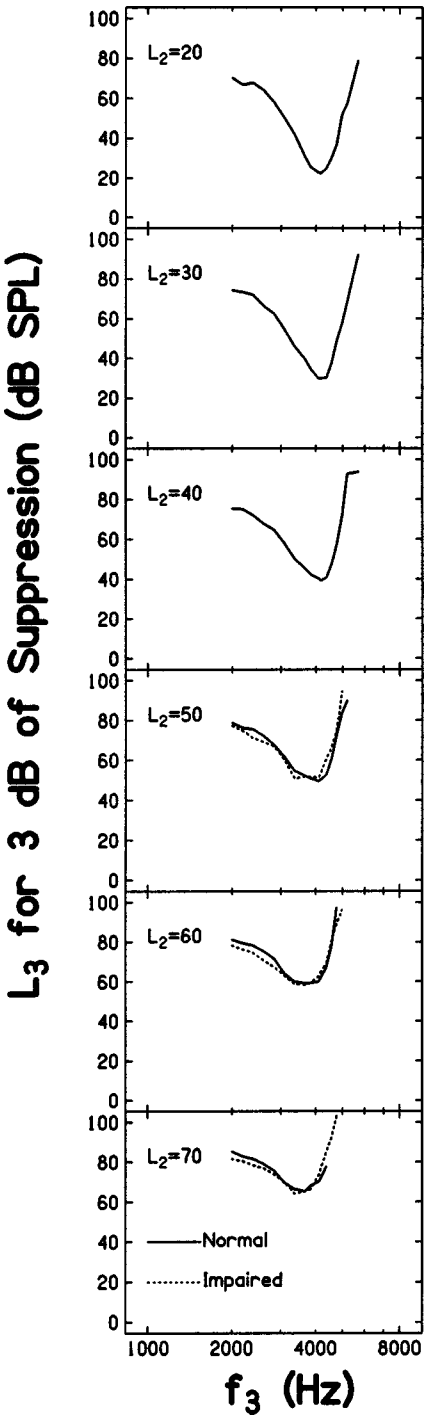


FIG. 7. Mean DPOAE STCs for normal-hearing and hearing-impaired ears plotted as the suppressor level in dB SPL for 3 dB of suppression as a function of f_3. Each panel shows the STC for a different L_2. STCs for subjects with normal hearing are shown as solid lines, while STCs for subjects with hearing loss are shown as dotted lines.

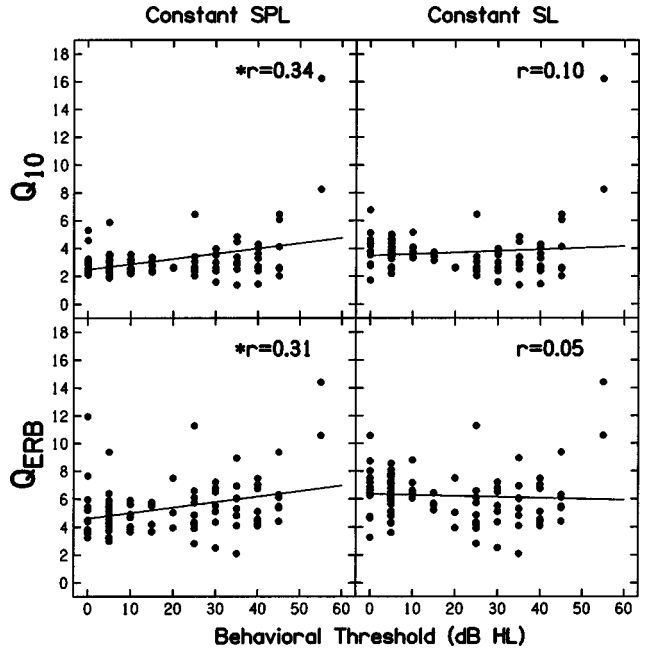


FIG. 8. Scatter plots of Q_{10} (top row) and Q_{ERB} (bottom row) as a function of audiometric threshold (dB HL) for the case when stimuli were presented at a constant SPL (left column) or constant SL (right column).

were inverted for the purpose of computing the ERB. While Q_{10} is the more common way of quantifying the sharpness of both neural and DPOAE tuning curves, Q_{ERB} has been used to estimate the tuning in behavioral estimates of frequency resolution (e.g., Shera, Guinan, and Oxenham, 2002).

The left column shows data from the two groups at constant SPL, but lumps together the data for the three L_2 levels summarized in each panel. Thus, it includes data from normal and impaired ears when $L_2 = 50$, 60, and 70 dB SPL. The right column similarly collapses data across three L_2 levels, but includes data when L_2 was more similar in SL. Thus, it includes data from normal ears when $L_2 = 20-40$ dB SPL, but includes data from impaired ears when $L_2 = 50-70$ dB SPL. As can be seen in Fig. 8, there was considerable overlap between normal and impaired *Q*s, and the relation between these values and audiometric thresholds was not obvious. This was true, whether the relation was evaluated at equivalent SPL or equivalent SL. Correlation analyses were consistent with this observation at equivalent SL (right column, Fig. 8). However, correlations ($r = 0.34$ or 0.31) were significant for both Q_{10} and Q_{ERB} at equivalent SPL (left column, Fig. 8). In spite of these correlations, an examination of the actual data suggests that a strong relationship between tuning around the tip of the STC and audiometric threshold did not exist, at least for the data from the present group of subjects.

Figure 9 shows mean Q_{10} and Q_{ERB} in the top and bottom panels, respectively, as a function of L_2 for both normal and impaired ears. Regardless of which estimate was used, the sharpness of tuning decreased as L_2 increased for subjects with normal hearing. Mean Q_{10} values of about 4 were observed when L_2 was 20 or 30 dB SPL, decreasing to a value of 2.6 at the highest probe levels. Similarly, Q_{ERB} decreased from average values of 6.5 at low L_2 levels to about 4.4 for L_2 levels of 60 dB SPL or greater. For ears with hearing loss, Q_{10} was slightly larger, compared to values observed in ears with normal hearing, when the estimates were derived at the same absolute L_2 levels. Q_{ERB} estimates also were slightly larger in ears with hearing loss. For both Q_{10} and Q_{ERB}, the differences between groups were small; however, an ANOVA for constant SPL conditions revealed that these differences, although small, were significant. No significant differences were noted when the data for normal

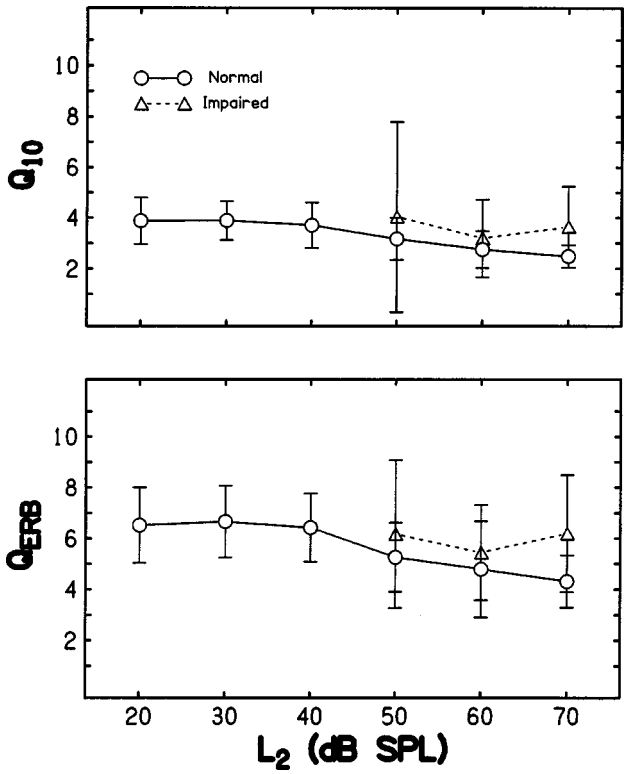


FIG. 9. Mean Q_{10} and Q_{ERB} as a function of L_2 in the top and bottom panels, respectively. Error bars represent ± 1 s.d. Data for subjects with normal hearing are shown as circles, while data for ears with hearing loss are shown as triangles.

ears at 20–40 dB SPL were compared to data from impaired ears at 50–70 dB SPL, which is the equivalent SL condition. Whether comparisons were made at constant SPL or SL, and in spite of the statistical observations, these estimates were variable in both normal and impaired ears, and the distributions of normal and impaired *Q*s overlapped, as can be seen in plots of individual (Fig. 8) and mean (Fig. 9) data.

F. Tip-to-tail differences (dB)

Following an approach initially proposed by Mills (1998), based on data from animal studies, and since used by Pienkowski and Kunov (2001) and Gorga *et al.* (2002b) to describe similar human DPOAE data, suppression thresholds were compared for a suppressor at the STC tip (i.e., close to f_2) and for a suppressor on the low-frequency tail of the STC. The differences between these suppression thresholds were estimated for the range of L_2 levels at which the measurements were made in each subject group. This quantity, which is specified in dB, is sometimes referred to as the tip-to-tail difference or the tip-to-tail ratio. The tails of the STCs were not completely flat, meaning that suppression threshold continued to slowly increase as f_3 decreased. As a consequence, the "tail" threshold was defined as the threshold for the 2.2–kHz suppressor because it was close to the lowest f_3 used in the present experiment and it was a suppressor frequency for which criterion suppression was measurable in the majority of cases. The tip-to-tail differences were derived for normal and impaired ears at all L_2 levels at which STCs could be measured.

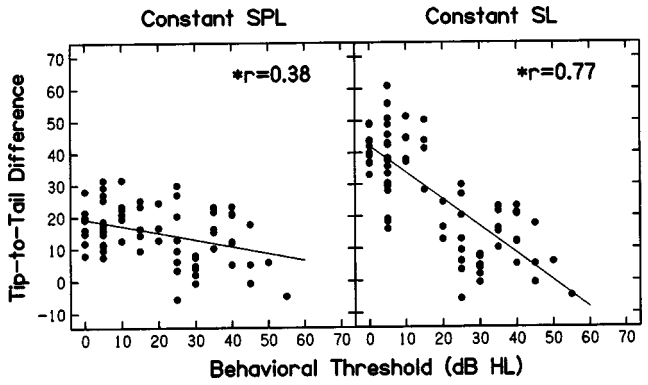


FIG. 10. Scatter plots of tip-to-tail difference (dB) as a function for audiometric threshold (dB HL). Left panel: data for stimuli presented at a constant SPL. Right panel: data for stimuli presented at a constant SL.

Figure 10 provides scatter plots of individual tip-to-tail differences as a function of audiometric threshold, following the convention that was used in Fig. 8. Thus, data are presented for equivalent SPL conditions (left panel) and equivalent SL conditions (right panel). Data are collapsed across the three L_2 levels represented in each panel, just as they were for the Q values represented in Fig. 8. A significant relationship was observed between tip-to-tail differences and audiometric thresholds, in that these differences decreased as audiometric thresholds increased, even though the data were variable at each audiometric threshold. Correlations ranged from 0.38 for the equivalent SPL condition to 0.77 for the equivalent SL condition. This represents a main observation in the present study. While tuning around the tip of the STCs did not depend strongly on audiometric threshold, the relationship between threshold at the tip and on the tail of the STC decreased systematically as threshold increased.

The results shown in Fig. 10 are summarized in Fig. 11, where the top panel plots the mean suppressor level (L_3) necessary for 3 dB of suppression as a function of L_2 for subjects with normal hearing. The parameter is f_3, with filled circles showing data for the case when $f_3 \approx f_2$ and open circles showing the results when $f_3 < f_2$. Note that when $f_3 \approx f_2$, the suppressor level necessary to achieve 3 dB of suppression was nearly equal to probe level (L_2), which also means that criterion suppressor level increased linearly with L_2. In contrast, a higher L_3 was required for 3 dB of suppression when the suppressor was 2.2 kHz, as expected from the individual STCs shown in Fig. 6 and the scatter plots in Fig. 10. For example, when $L_2 = 20$ dB SPL, an L_3 of about 65 dB SPL was needed in order for the criterion suppression to occur. However, L_3 increased at a slower rate, compared to L_2, for this low-frequency suppressor.

In similar fashion, the middle panel plots mean suppression thresholds as a function of L_2 for the same on-frequency and low-frequency suppressors in subjects with hearing loss. In this case, filled triangles represent results when $f_3 \approx f_2$, and open triangles represent data when $f_3 = 2.2$ kHz. As expected from previous results, suppression was measurable over a restricted range of L_2 levels in impaired ears, compared to ears with normal hearing, and data from impaired ears were characterized by greater variability. However, for those L_2 levels at which suppression could be measured, the

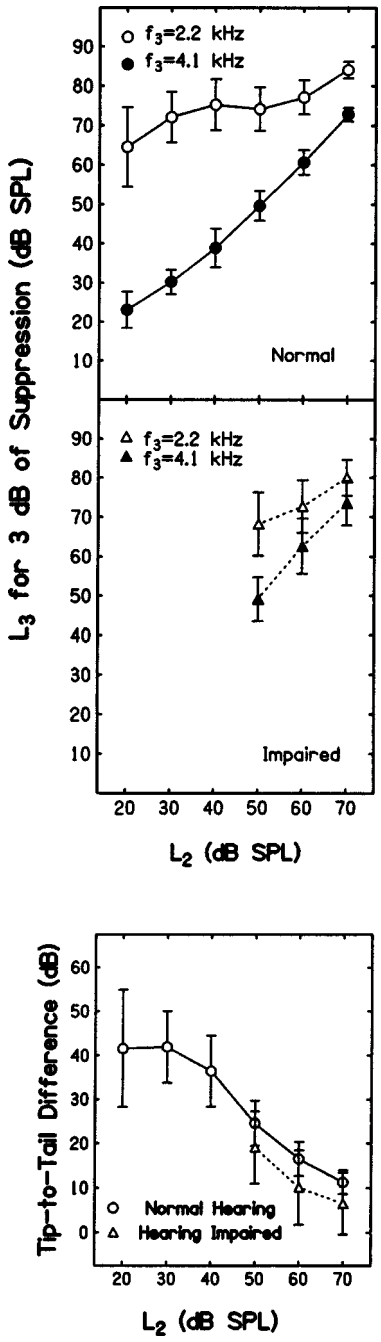


FIG. 11. Mean suppressor level (L_3) necessary for 3 dB of suppression (in dB SPL) as a function of probe level (L_2) for an on-frequency suppressor ($f_3=4.1$ kHz, filled symbols) and for a low-frequency suppressor ($f_3=2.2$ kHz, open symbols). Error bars represent ± 1 s.d. Top panel presents data from normal-hearing subjects; middle panel shows results for subjects with hearing loss. Bottom panel: The tip-to-tail difference (in dB as a function of L_2; data from subjects with normal hearing are shown as circles, while data from subjects with hearing loss are shown as triangles.

overall pattern of response for low- and on-frequency suppressors appears to be similar for the two groups. That is, the ''threshold'' suppressor level approximated L_2 and, thus, grew linearly with L_2 when $f_3 \approx f_2$. Higher suppressor levels were needed when $f_3=2.2$ kHz, while L_3 increased more slowly for this condition, compared to the case when $f_3 \approx f_2$.

The data shown in the top two panels are summarized in the bottom panel of Fig. 11, in which the mean levels for suppression thresholds when $f_3=4.1$ kHz were subtracted from the thresholds when $f_3=2.2$ kHz, and this quantity is plotted as a function of L_2. Thus, the bottom panel plots the mean tip-to-tail difference in dB for both subject groups. In normal-hearing ears, the tip-to-tail difference decreased as L_2 increased, going from a maximum of about 42 dB ($L_2=20$ dB SPL) to a minimum of about 12 dB ($L_2=70$ dB SPL). At the three L_2 levels at which data were obtained in impaired ears, tip-to-tail differences also systematically decreased with L_2. Ears with hearing loss, however, produced tip-to-tail differences that were 5–6 dB smaller than those produced by subjects with normal hearing at the same L_2 levels, although variability in these measurements was large, especially among ears with hearing loss. An ANOVA revealed significant effects for both L_2 and group, but no interactions between L_2 and group. This was true for both the constant SPL and the constant SL conditions.

IV. DISCUSSION

To summarize the results from the present experiment, threshold and growth of suppression (measured in the form of DPOAE decrement vs L_3 functions) depended on the relationship between suppressor frequency (f_3) and probe frequency (f_2) both in normal-hearing and hearing-impaired subjects (Figs. 3, 4, and 5). When $f_3<f_2$, suppression threshold occurred at higher levels and the slope of the decrement function was steeper, compared to cases when $f_3 \approx f_2$. While suppression threshold increased as f_3 increased above f_2, the slopes of the decrement functions decreased in both subject groups. There were differences, however, between normal and impaired ears in that the differences in slope between low-frequency ($f_3<f_2$) and high-frequency suppressors ($f_3>f_2$) were less in impaired ears. The decrement-vs-L_3 data were used to generate DPOAE STCs, in which the suppressor level necessary for 3 dB of suppression was plotted as a function of f_3 (Figs. 6 and 7). While STCs were measurable over a more limited range of absolute probe levels (L_2) in ears with hearing loss, at the same absolute probe levels, mean STCs in normal and impaired ears were similar in appearance. These similarities were noted around the tips of the STCs, in that there was overlap in the estimates of Q_{10} and Q_{ERB} for the two groups (Figs. 8 and 9). There was a tendency towards sharper tuning around the tip of the STC in ears with hearing loss, although this effect was significant only for conditions of equivalent absolute L_2 levels in normal and impaired ears. In contrast, ears with hearing loss produced smaller tip-to-tail differences, compared to subjects with normal hearing, with a larger effect for comparisons in which L_2 was presented at roughly equivalent SL (Figs. 10 and 11).

The relationship between the slopes of decrement functions as a function of suppressor frequency in normal ears has been observed in several previous DPOAE studies (e.g., Kemp and Brown, 1983; Kummer *et al.*, 1995; Abdala, 1998, 2001; Gorga *et al.*, 2002b). In turn, these data are consistent with previous animal studies in which it was shown that the response at any point along the cochlea grows more rapidly when that place is driven by a stimulus whose frequency is

lower than the best frequency for that place (Rhode, 1971; Sachs and Abbas, 1974; Schmiedt and Zwislocki, 1980; Ruggero and Rich, 1991; Ruggero *et al.*, 1997).

The observation that the slopes of these functions in ears with mild-moderate hearing loss were, in general, shallower for low-frequency suppressors, compared to data in normal hearing ears was not expected. Previous studies in animals suggest that response growth changes as a consequence of cochlear damage (Evans, 1974; Sewell, 1984; Gorga and Abbas, 1981a,b; Ruggero and Rich, 1991). A recent study in animals with noise-induced hearing loss described best-frequency rate-level functions that sometimes were shallower than normal, normal, or steeper than normal (Heinz *et al.*, 2003). They attributed the variation in slopes to differences in cochlear damage (IHC damage, OHC damage, or some combination of IHC and OHC damage), although cochlear status was not assessed directly. Likewise, the status of the OHCs and the IHCs cannot be known in the present study. In those studies that examined response growth for on- and off-frequency stimuli, however, little or no change in response growth was observed following cochlear damage for a stimulus that was lower than the best or probe frequency. Stated differently, the response to a low-frequency stimulus was relatively insensitive to the status of the cochlea at a high-frequency place. However, the slope of response-growth functions tended to increase (relative to the normal case) for stimuli close to the best frequency when cochlear damage existed. The present results appear, at least at face value, to differ from these previous findings. There was more similarity in the slope of these decrement functions for low- and high-frequency suppressors in impaired ears. This occurred because there was a small decrease in slope for suppressors below f_2 and perhaps an even smaller increase in slope for suppressors above f_2.

The differences between the present findings and predictions from studies in lower animals may be due to the differences in the measurement paradigms used to collect the data. Unlike previous animal experiments, in which response growth can be examined directly by presenting a single tone, such measurements are not possible in humans. In studies involving humans, a probe (the primaries in the present case) must be presented and response growth to another tone (the suppressor) must be inferred from the changes it causes to the probe response. The integrity of the OHCs presumably should not influence the response to a low frequency at a high-frequency place because the high-frequency place only responds nonlinearly when it is excited by frequencies close to its best frequency. Evidence in support of this view comes from single-unit studies in animals where OHC damage exists. Tail thresholds remain relatively constant (or perhaps even hypersensitive) so long as the damage is restricted to the OHCs (Dallos and Harris, 1978; Liberman and Dodds, 1984). Further support for this view comes from studies in which basilar-membrane motion was measured before and after treatment with furosemide, an agent known to reversibly impair cochlear function (Ruggero and Rich, 1991). While the response from a high-frequency cochlear place to best-frequency tones changed when furosemide was administered, there was little change at that same place when it was stimulated with a low-frequency tone.

It may be reasonable to assume, therefore, that the effects of the low-frequency suppressors were the same in normal and impaired ears. However, the response to suppressors must be inferred from the changes their presentation causes to the response to the probe. Perhaps the slope patterns observed in the present study are a consequence of the fact that our measures reflect the relative growth of response between suppressor and probe. If the growth of response was more alike in impaired ears for low- and on-frequency conditions, then differences in slope across frequency might be reduced, not because the low-frequency slope has become more shallow, but because the on-frequency slope has become more steep.

Measures of tuning, based on the present measurements, also revealed some interesting trends. DPOAE STCs in normal and mildly impaired ears appeared to be similar when they were compared at the same probe levels (L_2). In spite of the statistical outcome indicating that impaired ears produced larger Q_{10} and Q_{ERB} than normal-hearing ears when the probes were presented at equivalent SPL, it is difficult to consider the individual data presented in Fig. 8 and conclude that these differences are meaningful. That is, there was overlap between Q_{10} and Q_{ERB}, regardless of audiometric threshold. These results, at one level, are similar to those observed in two recent papers in which either loop diuretics (Martin *et al.*, 1998) or noise exposure (Howard *et al.*, 2002) was used to induce temporary, reversible effects in the cochleae of rabbits. This group recently observed similar effects when more permanent noise-exposure damage was induced in rabbits (Howard *et al.*, 2003). In all three studies, there was little difference in the DPOAE STCs prior to and soon after cochlear insult (although some STCs in impaired ears appeared more variable compared to the pre-exposure or pre-treatment cases). In fact, there may actually have been a tendency for some STCs reported in these previous studies to become more sharply tuned after insult. In any case, Martin *et al.* and Howard *et al.* observed little or no change in either Q or tip-to-tail differences following cochlear insult. As a result of these findings, they concluded that DPOAE suppression measurements do not provide the same information about cochlear tuning as do other measurements, such as single-unit frequency-threshold curves or direct measurements of basilar-membrane tuning. The present observation of little or no difference between the DPOAE STC sharpness around the tip (Q) in normal and impaired ears would be consistent with the observations made by Martin *et al.* and Howard *et al.* There are other physiological data in the form of single-unit frequency-threshold curves (Dallos and Harris, 1978; Liberman and Dodds, 1984) or forward-masking action potential (AP) tuning curves (Gorga and Abbas, 1981b), suggesting that tuning around the best frequency (single-unit studies) or probe frequency (AP studies) may be similar in ears with normal and impaired hearing, at least for ears with mild-to-moderate threshold elevations. Stated differently, hearing loss (by definition) causes an elevation of threshold at best frequency, but not necessarily a decrease in tuning around best frequency. Thus, the present observations in re-

gard to Q are consistent with previous DPOAE data and with the results of single-unit and AP masking studies in animals with induced lesions. While the present findings in relation to tip-to-tail differences are also in agreement with previous single-unit and AP data, they differ from DPOAE data in animals with induced cochlear damage that showed no change in tip-to-tail differences (Martin *et al.*, 1998; Howard *et al.*, 2002, 2003).

Conclusions about tuning at the tip in normal and impaired ears might be modified by examining level effects. In the present data, Q_{10} and Q_{ERB} (in addition to tip-to-tail differences) decreased as level increased, an observation that was more pronounced in ears with normal hearing. Larger Qs were seen at low probe levels (L_2 levels of 20–40 dB SPL), compared to L_2 levels at and above 50 dB SPL. Less systematic changes with level were observed in impaired ears. However, this may have occurred because of the restricted range of possible measurements in impaired ears. Hearing loss in these subjects increased DPOAE threshold and/or reduced DPOAE level (see Fig. 1). Because suppression experiments require an unsuppressed response of some level that subsequently will be reduced by the suppressor, it was necessary to perform the experiments in impaired ears only at the higher L_2 levels for which some reduction in tuning occurs in normal ears. This suggests that the lack of differences between Qs in normal and impaired ears might represent a level effect. Thus, the observation of differences between normal and impaired ears might depend on the way the level is chosen at which comparisons will be made. This argument, however, is not supported by the observation that the overlap in Qs was similar, regardless of whether normal and impaired data were compared at equivalent SPL or equivalent SL (see Fig. 8). The present data, therefore, suggest that Q is relatively insensitive to probe level (L_2) or audiometric threshold, at least among subjects with no worse than a moderate hearing loss, statistical results for Q notwithstanding.

The present estimates of Q_{ERB} in human ears with normal hearing are less than the values recently reported by Shera *et al.* (2002), whose estimates were based on stimulus-frequency otoacoustic emission (SFOAE) and behavioral forward-masking measurements in humans. For example, their SFOAE estimates of Q_{ERB} were 15–20 at 4 kHz, while their behavioral Q_{ERB} estimates were close to 15 at 4 kHz. Their measurements were made with a 40-dB SPL probe. At this level, the current estimates of Q_{ERB}, on average, were about 6.5. This value is closer to the values (about 8 or 9 at 4 kHz) observed in a number of other papers (as summarized by Glasberg and Moore, 1990), in which simultaneous masking techniques were used. Thus, the differences between our estimates and the recent estimates by Shera *et al.* might relate to the fact that the data reported by Shera *et al.* represent the excitatory response area, while the present measurements (as well as those summarized by Glasberg and Moore) include suppressive, as well as excitatory, areas in the response. The SFOAE measurements of Shera *et al.*, although derived in a suppression paradigm, represent the unsuppressed response to a single tone. Similarly, their forward-masking behavioral data presumably describe only the excitatory representation to the probe and not the wider representation due to suppression (see Sachs and Kiang, 1968, for a classic description of excitatory and suppressive regions in the responses of auditory neurons). This would be the case because suppression is memoryless, occurring only when the probe and suppressor (or masker) are on at the same time. In contrast to the paradigm used by Shera *et al.*, the present measurements were derived from suppression measurements, and thus would be expected to result in wider estimated bandwidths. In addition, the probe in the present study consisted of two tones, slightly different in frequency. Both tones contribute to the generation of the DPOAE, and, because of this, it is possible that a wider range of suppressors is effective in reducing the response. While both factors may account for the differences in Q estimates between the present study and the data reported by Shera *et al.*, it is uncertain how this distinction would impact the present comparisons between normal and impaired ears. Perhaps an effect of hearing status would have been observed (in Q_{10} and Q_{ERB}) if the data were derived from a paradigm that did not include the effects of suppression. A paradigm in which only the excitatory regions were evaluated might have resulted in sharper tuning (higher Q) in normal ears because the suppression region would not have been outlined, but might have had no effect on the tuning in impaired ears. Unfortunately, isolating the excitatory region in normal ears is not possible for measurements like those used in the present study.

In normal ears, it is assumed that nonlinear processing occurs at a specific place when that place is driven by its best or characteristic frequency. In suppression experiments like the present study or psychophysical masking studies (Oxenham and Plack, 1997), the growth of suppression (or masking) with probe level is linear when probe frequency approximates suppressor frequency because both the suppressor (or masker) and the probe are being processed through the same nonlinear mechanism. In contrast, low-frequency suppressors (or maskers) are processed more linearly at the place where the probe is primarily represented. Thus, a high-frequency probe is processed nonlinearly (compressively) at its characteristic place, while a low-frequency suppressor is processed linearly at the same place. As a consequence, functions relating suppressor level to probe level grow at a slow rate when suppressor frequency is well below probe frequency, reflecting the compressive processing for the probe, but not for the low-frequency suppressor (see the top two panels of Fig. 11).

These effects were explored further in the present study by estimating tip-to-tail differences (Figs. 10 and 11). Tip-to-tail differences decreased as either level or audiometric threshold increased. When grouped dichotomously according to hearing status, normal and impaired results differed significantly, regardless of whether comparisons were made at the same SPL or the same SL. These data suggest that one effect of hearing loss might be to cause more upward spread of excitation in ears with hearing loss, even if the tuning close to the best frequency (f_2 in the present experiment) does not differ between normal and impaired ears. These results are consistent with single-unit data (e.g., Dallos and

Harris, 1978; Liberman and Dodds, 1984), in which the differences in tip and tail sensitivity decreased as the extent of the cochlear lesion (and the amount of threshold elevation) increased. In addition, they are consistent with forward-masking AP data, in which tip-to-tail differences in normal ears were larger than those observed in ears with noise-induced hearing loss, regardless of whether the probe in the normal ears was presented at the same SL or the same SPL as was used in impaired ears (Gorga and Abbas, 1981b).

The influence of stimulus level appears to be consistent with data reported by Rubsamen *et al.* (1995), in which DPOAE levels and neuronal thresholds were correlated before and after administration of furosemide. At low and moderate primary levels, the DPOAE level decreased and the neuronal threshold increased synchronously. Decreases in DPOAE level of 30 dB or more and increases in unit thresholds of 50 dB or more occurred at essentially the same time. As the primary levels increased, however, the size of the reduction in DPOAE level decreased. At primary levels of 80 and 70 dB SPL (L_1 and L_2, respectively), smaller changes in DPOAE level were observed (see Fig. 5, Rubsamen *et al.*, 1995, for an example of these effects). Stated differently, changes in DPOAE level were more apparent when low-level primaries were used to elicit these responses, compared to the case when higher-level stimuli were used. This effect is consistent with other reports that have shown that the sensitivity of DPOAE measurements to hearing loss is greater for lower level primaries (e.g., Whitehead *et al.*, 1995; Stover *et al.*, 1996).

Mills (1998) described an approach in which DPOAE STCs were used to calculate differences between on-frequency and low-frequency thresholds, and proposed that these differences related to the ''gain of the cochlear amplifier.'' As stated above, this idea is based on the notion that active nonlinear processes occur when a given place in the cochlea is driven by its ''best'' or ''characteristic'' frequency (CF), and these nonlinear processes are absent when the same place is driven by a frequency much lower than CF. Pienkowski and Kunov (2001) applied this approach with humans whose audiometric thresholds fell within broad normal limits, using a single set of moderate-level primaries. We further explored this application by measuring tip-to-tail differences for a wide range of primary levels in humans with normal hearing (Gorga *et al.*, 2002b). The tip-to-tail differences previously observed by us decreased with primary levels in a manner that was similar to what was observed for normal-hearing subjects in the present study. One interpretation of these data might be that the gain of the cochlear amplifier (if that is what is being estimated by measurements of tip-to-tail differences) decreases as level increases. It is as if the need for amplification of low-level stimuli decreases as level increases. If this notion is correct, it might provide a framework for interpreting the present results. Specifically, both hearing status and probe level might be exerting an influence on the tip-to-tail difference in impaired ears. The ''gain'' in ears with hearing loss could only be measured for L_2 levels of 50–70 dB SPL, for the reasons described previously. These are levels for which, under normal circumstances, less gain is evident. Thus, differences between normal and impaired ears are reduced when compared at the same absolute levels because of changes with level to the nonlinear processing in ears with normal hearing. Added to this level effect is the effect of hearing loss. Thus, the small differences between normal and impaired ears (at equivalent SPL) relate to the fact that comparisons could be made only for high-level responses, with their attendant reduced gain under normal conditions. The further reduction in the tip-to-tail differences in ears with hearing loss might reflect an additional consequence of hearing loss, namely an increase in upward spread of excitation.

There are several potentially important differences between the way single-unit frequency-threshold curves (FTC) and psychophysical tuning curves (PTC) are measured, compared to how the present DPOAE STCs were derived. In the case of single-unit FTCs, ''threshold'' typically is defined as the stimulus level that results in a specified increase in discharge rate above the spontaneous rate (for example, see Liberman, 1978). Thus, the threshold represents a level that causes a small response on the nerve fiber. PTCs typically are measured by setting a probe tone slightly above its threshold, and then presenting maskers that render this barely audible sound inaudible. In contrast, DPOAE STCs typically are measured by presenting the probe (f_2 and f_1) at a level that produces a response well above its threshold. This suprathreshold response is then suppressed by some criterion amount (i.e., decrement) in order to produce a DPOAE STC. It may be an important distinction that both FTCs and PTCs are measured at near-threshold levels (either as an increase in discharge rate or as the masking of a low-level probe), while DPOAE STCs are measured as reductions in level for a suprathreshold response that typically is elicited by moderate level probes. Howard *et al.* (2002) made this point when noting the unexpected finding that DPOAE STCs in ears exposed to noise did not show changes in tuning-curve properties that might be predicted from FTCs measured at the level of an auditory neuron. Perhaps more similarity would be observed if DPOAE STCs were measured in paradigms more skin to the paradigms used in FTC or PTC measurements.

Another important difference relates to the levels in the auditory system that are assessed during DPOAE measurements, compared to either single-unit or psychophysical studies. It is reasonable to assume that DPOAE studies describe only OHC function, and are uninfluenced by the status of the IHCs, auditory nerve, or higher levels of the auditory system (admittedly ignoring influences from the efferent system). In contrast, single-unit studies provide information related to both OHC and IHC status, and psychophysical masking studies (although dominated by the status of the auditory periphery) also may be affected by more central phenomena.

Finally, there is perhaps a parsimonious (and less interesting) explanation for the present findings, related to subject selection criteria. Special effort was made to recruit subjects with mild-to-moderate hearing loss at 4 kHz. However, this inclusion criterion alone was insufficient. It was necessary that each ear produce a DPOAE of sufficient level in order to conduct these suppression studies. A total of 29 ears of 25

subjects with mild-to-moderate hearing loss was recruited, but DPOAE suppression experiments were possible in only 23 ears of 21 of these subjects. These subjects produced responses of sufficient level to permit measurements of suppression of that response. Thus, the subjects with hearing loss included in the study constitute a biased sample in that they produced larger DPOAEs than the excluded subjects. While this was necessary in order to conduct suppression experiments, it suggests the possibility that the present subjects are NOT entirely representative of patients with mild-to-moderate hearing loss. This caveat regarding bias in the selection of subjects with hearing loss may be unnecessary, given the fact that the majority of ears with mild-to-moderate hearing loss produced responses.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH (NIDCD R01 DC02251). Portions of this work were presented at the 2003 Midwinter Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL. The recruitment of subjects and the development of some of the software were supported, in part, by a Core Grant (P30 DC04662). We thank Sandy Estee for her assistance in subject recruitment. We also thank Andrea Pittman for her help in data analyses and in the preparation of some figures. Finally, we would like to thank the two anonymous reviewers who encouraged us to perform statistical analyses on these data.

Abdala, C. (**1998**). ''A developmental study of distortion product otoacoustic emission ($2f_1-f_2$) suppression in humans,'' Hear. Res. **21**, 125–138.

Abdala, C. (**2001**). ''Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary levels,'' J. Acoust. Soc. Am. **110**, 1465–1476.

Abdala, C., Sininger, Y. S., Ekelid, M., and Zeng, F.-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for Audiometers'' (American National Standards Institute, New York).

Boege, P., and Janssen, T. (**2002**). ''Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O functions in normal and cochlear hearing loss ears,'' J. Acoust. Soc. Am. **111**, 1810–1818.

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f_1-f_2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Dallos, P. J., and Harris, D. M. (**1978**). ''Properties of auditory-nerve responses in the absence of outer hair cells,'' J. Neurophysiol. **41**, 365–383.

Dallos, P. J., Harris, D. M., Relkin, E., and Cheatham, M. A. (**1980**). ''Two-tone suppression and intermodulation distortion in the cochlear: Effect of outer hair cell lesions,'' in *Psychophysical, Physiological and Behavioral Studies of Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft, The Netherlands), pp. 242–252.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). ''Distortion-product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears,'' J. Acoust. Soc. Am. **110**, 3119–3131.

Evans, E. F. (**1974**). ''Auditory frequency selectivity and the cochlear nerve,'' in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, New York), pp. 118–129.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Gorga, M. P., and Abbas, P. J. (**1981a**). ''AP measurements of short-term adaptation in normal and in acoustically traumatized ears,'' J. Acoust. Soc. Am. **70**, 1310–1321.

Gorga, M. P., and Abbas, P. J. (**1981b**). ''Forward-masking AP tuning curves in normal and otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 968–977.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). ''Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses,'' J. Acoust. Soc. Am. **93**, 2050–2060.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (**2003**). ''Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions,'' J. Acoust. Soc. Am. (in press).

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). ''From laboratory to clinic: A large-scale study of distortion product otoacoustic emissions in ears with normal and ears with hearing loss,'' Ear Hear. **18**, 440–455.

Gorga, M. P., Nelson, K., Davis, T., Dorn, P. A., and Neely, S. T. (**2000**). ''Distortion product otoacoustic emission test performance when both $2f_1-f_2$ and $2f_2-f_1$ are used to predict auditory status,'' J. Acoust. Soc. Am. **107**, 2128–2135.

Gorga, M. P., Stover, L. J., and Neely, S. T. (**1996**). ''The use of cumulative distributions to determine critical values and levels of confidence for clinical distortion product otoacoutic emission measurements,'' J. Acoust. Soc. Am. **100**, 968–977.

Gorga, M. P., Neely, S. T., and Dorn, P. A. (**2002a**). ''Distortion product otoacoustic emissions in relation to hearing loss,'' in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 243–272.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Konrad-Martin, D. (**2002b**). ''The use of distortion product ototacoustic emission suppression as an estimate of response growth,'' J. Acoust. Soc. Am. **111**, 271–284.

Heinz, M. G., Sachs, M. B., and Young, E. D. (**2003**). ''Activity growth rates in auditory-nerve fibers following noise-induced hearing loss,'' Twenty-Sixth Annual Midwinter Research Meeting of the Association for Research in Otolaryngology.

Howard, M. A., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2002**). ''Effects of reversible noise exposure on the suppression tuning of rabbit distortion-product otoacoustic emissions,'' J. Acoust. Soc. Am. **111**, 285–296.

Howard, M. A., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2003**). ''Suppression tuning of rabbit distortion-product otoacoustic emissions following permanently damaging acoustic overexposure,'' Twenty-Sixth Annual Midwinter Research Meeting of the Association for Research in Otolaryngology.

Kemp, D. T., and Brown, A. M. (**1983**). ''A comparison of mechanical nonlinearities in the cochleae of man and gerbil from ear canal measurements,'' in *Hearing: Physiological Basis and Psychophysics*, edited By R. Klinke and R. Hartman (Springer, Berlin), pp. 82–88.

Kiang, N. Y.-S., 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.

Kim, D. O. (**1980**). ''Cochlear mechanics: Implications of electrophysiological and acoustical observations,'' Hear. Res. **2**, 297–317.

Kim, D. O., Paparello, J., Jung, M. D., Smursynski, J., and Sun, X. (**1996**). ''Distortion product otoacoustic emission test of sensorineural hearing loss: Performance regarding sensitivity, specificity, and receiver operating characteristics,'' Acta Otolaryngol. (Stockh) **116**, 3–11.

Kummer, P., Janssen, 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.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.

Liberman, M. C. (**1978**). ''Auditory-nerve responses from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.

Liberman, M. C., Gao, J., He, D. Z., Wu, X., Jia, S., and Zuo, J. (**2002**). ''Prestin is required for electromotility of the outer hair cells and for the cochlear amplifier,'' Nature (London) **419**, 300–304.

Liberman, M. C., and Dodds, L. W. (**1984**). ''Single-unit labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves,'' Hear. Res. **16**, 55–74.

Martin, G. K., Ohlms, L. A., Franklin, D. J., Harris, F. P., and Lonsbury-Martin, B. L. (**1990**). ''Distortion product otoacoustic emissions in humans. III. Influence of sensorineural hearing loss,'' Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 30–42.

Martin, G. K., Stagner, B. B., Jassir, D., Telischi, F. F., and Lonsbury-Martin, B. L. (**1999**). ''Suppression and enhancement of distortion-product

otoacoustic emissions by interference tones above f_2. I. Basic findings in rabbits,'' Hear. Res. **136**, 105–123.
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,'' Hear. Res. **28**, 191–208.
Martin, G. K., Jassir, D., Stagner, B. B., and Lonsbury-Martin, B. L. (**1998**). ''Effects of loop diuretics on the suppression tuning of distortion-product otoacoustic emissions in rabbits,'' J. Acoust. Soc. Am. **104**, 972–983.
Mills, D. M. (**1998**). ''Interpretation of distortion product otoacoustic emission measurements. II. Estimating tuning characteristics using three stimulus tones,'' J. Acoust. Soc. Am. **103**, 507–523.
Neely, S. T., and Gorga, M. P. (**1998**). ''Comparison between intensity and pressure as measures of sound level in the ear canal,'' J. Acoust. Soc. Am. **104**, 2925–2934.
Neely, S. T., and Liu, Z. (**1994**). ''EMAV: Otoacoustic emission averager,'' Tech. Memo No. 17 (Boys Town National Research Hospital, Omaha, NE).
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Pienkowski, M., and Kunov, H. (**2001**). ''Suppression of distortion product otoacoustic emissions and hearing thresholds,'' J. Acoust. Soc. Am. **109**, 1496–1502.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkey using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rubsamen, R., Mills, D. M., and Rubel, E. W. (**1995**). ''Effects of furosemide on distortion product otoacoustic emissions and on neuronal responses in the anteroventral cochlear nucleus,'' J. Neurophysiol. **74**, 1628–1638.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of cortin mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neuro. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Sachs, M. B., and Abbas, P. J. (**1974**). ''Rate versus level functions for auditory-nerve fibers in cats: Tone-burst responses,'' J. Acoust. Soc. Am. **56**, 1835–1847.
Sachs, M. B., and Kiang, N. Y. (**1968**). ''Two-tone inhibition in auditory-nerve fibers,'' J. Acoust. Soc. Am. **43**, 1120–1128.
Schmiedt, R. A., and Zwislocki, J. J. (**1980**). ''Effects of hair cell lesions on responses of cochlear nerve fibers. II. Single- and two-tone intensity function in relation to tuning curves,'' J. Neurophysiol. **43**, 1390–1405.
Sewell, W. F. (**1984**). ''Furosemide selectively reduces one component in rate-level functions from auditory-nerve fibers,'' Hear. Res. **15**, 69–72.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Siegel, J. H. (**1994**). ''Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes,'' J. Acoust. Soc. Am. **95**, 2589–2597.
Siegel, J. H. (**2002**). ''Calibrating otoacoustic emission probes,'' in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 416–441.
Stover, L., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996**). ''Towards optimizing the clinical utility of distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 956–967.
Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Dependence of distortion product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L_2 below L_1,'' J. Acoust. Soc. Am. **97**, 2346–2358.

Suppression tuning in noise-exposed rabbits

MacKenzie A. Howard[a)]
Neuroscience Program, University of Miami School of Medicine, Miami, Florida 33101-6960 and Department of Otolaryngology, University of Colorado Health Sciences Center, Denver, Colorado 80262-0001

Barden B. Stagner
Department of Otolaryngology, University of Colorado Health Sciences Center, Denver, Colorado 80262-0001

Paul K. Foster
Department of Otolaryngology, University of Miami School of Medicine, Miami, Florida 33101-6960

Brenda L. Lonsbury-Martin and Glen K. Martin
Department of Otolaryngology, University of Colorado Health Sciences Center, Denver, Colorado 80262-0001

(Received 4 October 2002; accepted for publication 7 April 2003)

Psychophysical, basilar-membrane (BM), and single nerve-fiber tuning curves, as well as suppression of distortion-product otoacoustic emissions (DPOAEs), all give rise to frequency tuning patterns with stereotypical features. Similarities and differences between the behaviors of these tuning functions, both in normal conditions and following various cochlear insults, have been documented. While neural tuning curves (NTCs) and BM tuning curves behave similarly both before and after cochlear insults known to disrupt frequency selectivity, DPOAE suppression tuning curves (STCs) do not necessarily mirror these responses following either administration of ototoxins [Martin *et al.*, J. Acoust. Soc. Am. **104**, 972–983 (1998)] or exposure to temporarily damaging noise [Howard *et al.*, J. Acoust. Soc. Am. **111**, 285–296 (2002)]. However, changes in STC parameters may be predictive of other changes in cochlear function such as cochlear immaturity in neonatal humans [Abdala, Hear. Res. **121**, 125–138 (1998)]. To determine the effects of noise-induced permanent auditory dysfunction on STC parameters, rabbits were exposed to high-level noise that led to permanent reductions in DPOAE level, and comparisons between pre- and postexposure DPOAE levels and STCs were made. Statistical comparisons of pre- and postexposure STC values at CF revealed consistent basal shifts in the frequency region of greatest cochlear damage, whereas thresholds, $Q_{10\,dB}$, and tip-to-tail gain values were not reliably altered. Additionally, a large percentage of high-frequency lobes associated with third tone interference phenomena, that were exhibited in some data sets, were dramatically reduced following noise exposure. Thus, previously described areas of DPOAE interference above f_2 may also be studied using this type of experimental manipulation [Martin *et al.*, Hear. Res. **136**, 105–123 (1999); Mills, J. Acoust. Soc. Am. **107**, 2586–2602 (2002)].
[DOI: 10.1121/1.1577555]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Bt

I. INTRODUCTION

Aspects of frequency tuning have been described for many components of the auditory system. Typically, tuning curves are produced by plotting the stimulus level required to produce some criterion response as a function of varying stimulus frequency. Examples of such functions have been produced by recording single nerve-fiber responses to sound, i.e., neural tuning curves (NTCs), by measuring basilar-membrane (BM) amplitude or velocity, or by suppression of various types of otoacoustic emissions.

The first level at which the sharp frequency tuning characteristic of the auditory system can be measured is the BM (e.g., Robles *et al.*, 1986). When BM motion is stimulated using a single pure tone varied in level and frequency, plots of BM response (amplitude or velocity) as a function of frequency show that it is a highly tuned structure. These tuning curves have a steep high-frequency slope and a shallower low-frequency tail that often extends beyond half an octave below the most sensitive tip region. Thus, the auditory system is highly tuned at the BM level.

Electrophysiological recordings from single axons of the auditory nerve also exhibit responses with a high degree of frequency selectivity (Kiang *et al.*, 1965). Because each spiral ganglion neuron innervates a single inner hair cell, the responses of these axons reflect tuning from the intact organ of Corti (OC). NTCs are produced utilizing pure-tone stimuli identical to those used to measure BM tuning. Consequently, NTCs produced by plotting single nerve-fiber responses to acoustic stimuli across a wide frequency range exhibit tuning characteristics identical to BM tuning curves, i.e., steep high-

a)Electronic mail: mackenzie.howard@uchsc.edu

frequency slopes and shallow low-frequency tails. In fact, tuning curves of BM displacement and NTCs recorded from the same animal have been shown to exhibit the same characteristics in both normal (Narayan *et al.*, 1998) and damaged cochleas (Rhode, 1971).

A number of investigators (e.g., Brown and Kemp, 1984; Kummer *et al.*, 1995; Abdala *et al.*, 1996; Martin *et al.*, 1998b; Mills, 1998; Howard *et al.*, 2002) have also attempted to measure tuning of the auditory system using DPOAE interference techniques. In this procedure, the two primary tones, f_1 and f_2, necessary to elicit the DPOAE, are held constant, while a third interference tone (IT), f_3, is varied in frequency and level, to alter the response. Depending on the frequency and level of f_3, either suppression or enhancement of the DPOAE can result. STCs are produced by plotting the level of the IT required to reduce the DPOAE by a criterion amount, as a function of IT frequency. Regardless of the clear differences in stimulus parameters necessary to measure these STCs, the frequency tuning characteristics of the $2f_1-f_2$ DPOAE STC resemble the other measures of frequency tuning in the auditory system, described above.

Despite the similarities between the basic characteristics of STCs to other tuning measures, it is still not entirely clear how STCs relate to the more classic measures of auditory frequency selectivity represented by BM and neural tuning, or how they might be effective indicators of cochlear dysfunction. Three approaches have been taken in the literature to either directly or indirectly address these problems. The first approach involves direct comparison between NTCs and STCs. Such comparisons have been made in bats (Frank and Kossl, 1995) and lizards (Koppl and Manley, 1993). In each of these studies it is clear that STCs and NTCs share the common frequency tuning properties of sharp, sensitive tips at the best frequency, steep high-frequency slopes, and shallow low-frequency tails. The most striking example of the similarities between NTCs and STCs involves tuning curves recorded in the mustached bat. Over a relatively narrow frequency band, both STCs and NTCs exhibit attributes opposite of normal tuning curves (e.g., steep low-frequency slopes and shallow high-frequency tails). These changes in tuning properties coincide with a highly specialized region of the bat cochlea that is necessary for echolocation, and the abnormal tuning characteristics of the bat cochlea reflect these adaptations (Frank and Kossl, 1995).

A second major approach has been to monitor STC changes in cochlear function over the course of development. Several studies have recorded STCs in adult humans (Kummer *et al.*, 1995; Abdala *et al.*, 1996; Abdala, 1998; Pienkowski and Kunov, 2001), and interesting comparisons have been made between adult STCs and those recorded from neonatal humans (Abdala *et al.*, 1996). For example, it has been demonstrated that STCs from neonatal cochleae consistently have a sharper (i.e., have larger $Q_{10\,dB}$ values) shape, and it is thought that cochlear immaturity may be the cause of these alterations in suppression of the cochlear nonlinearity (Abdala, 1998). Documenting changes in DPOAE tuning is thus very important, as these changes may provide information about very specific types of changes in the functionality of the auditory system, without the invasive measurement of BM or neural tuning.

A third means of evaluating STCs involves manipulation of the auditory system to produce predictable changes in tuning based on past studies of BM or nerve-fiber physiology. Changes in frequency tuning, especially as measured by BM motion or collection of NTCs, have been well documented for a variety of different types of cochlear insult. Following damage, stimulus levels required to produce a criterion response are elevated, especially at the best frequency of the tuning curve, thus increasing the tip threshold. These increases in tip threshold result in a loss of tuning curve sharpness, which is reflected by decreases in the bandwidth of tuning measured by the $Q_{10\,dB}$ tuning curve parameter. DPOAE interference can also be monitored during such manipulations and following recovery of cochlear function.

One type of cochlear insult, whose effects on auditory frequency tuning have been well documented, is ototoxic poisoning with loop diuretics. The effects of chemicals such as furosemide on BM, neural, and DPOAE tuning have been described. Diuretic intoxication clearly disrupts the tuning of the BM, leading to reduced tuning-curve tip thresholds and decreased $Q_{10\,dB}$ measurements (Ruggero and Rich, 1991; Rubsamen *et al.*, 1995). Single nerve-fiber recordings show similarly altered frequency selectivity for the cochlear nerve (e.g., Evans and Klinke, 1982; Sewell, 1984). However, DPOAE STCs do not show such characteristic changes. The shape of these functions did not change appreciably in rabbits exposed to loop diuretics, even though tip thresholds exhibited consistent, but relatively small, elevations (Martin *et al.*, 1998b). Together, results from these studies show that diuretic administration disrupts BM and NTC frequency tuning, but not DPOAE tuning based on a criterion amount of suppression.

Similar comparisons have been made between different auditory tuning functions following other cochlear insults. Temporary noise-induced cochlear dysfunction is reflected by the deterioration of single cochlear nerve-fiber frequency selectivity (e.g., Kiang *et al.*, 1976; Salvi *et al.*, 1978; Liberman and Beil, 1979). As in the case of ototoxic poisoning, NTCs recorded following noise exposure show raised thresholds and decreased $Q_{10\,dB}$ values. As expected due to the direct link between BM and inner hair cell tuning, noise overexposure disrupts BM tuning in the same way (Johnstone *et al.*, 1986). Despite the clear alterations temporary noise-induced dysfunction has on these measures, STCs are affected quite differently by such exposures. Recently, it was shown that temporary noise-induced reductions in DPOAE level led to slight decreases in tuning curve tip threshold and increases in $Q_{10\,dB}$, resulting in functions with sharper tuning, rather than broader tuning (Howard *et al.*, 2002). Again, these differences can be taken as evidence that STCs do not directly measure BM or neural tuning when based upon a criterion amount of suppression, the traditional way these functions are generated.

The effects of permanent noise-induced damage on BM and neural tuning have also been documented (e.g., Salvi *et al.*, 1978, 1980; Liberman and Dodds, 1984; Walsh and McGee, 1990). Changes in these functions following exces-

sive noise exposures, from which the cochlea does not fully recover, are similar to behaviors described above for noise-induced temporary dysfunction. NTCs recorded from animals after such exposures show the stereotypically broad pattern of disrupted frequency selectivity. Generally, these changes are due to a loss of sensitivity to the best frequency, at which the tuning of nerve fibers is the most sharp. Changes in STCs following this type of noise exposure have not yet been documented. Due to the different levels of damage, and the different mechanisms by which cells in the inner ear are affected following temporarily and permanently damaging noise exposures, it cannot be assumed that STCs will behave in the same way during this type of dysfunction as they do during other sorts of cochlear insult.

The purpose of the present study was to provide a qualitative description of the effects of permanent noise-induced cochlear dysfunction on DPOAE properties and STCs in the rabbit. This work was a continuation of previous studies that documented changes in DPOAE suppression tuning following ototoxic insult and temporarily damaging noise. Standard DP-grams and STCs were recorded in rabbits prior to high-level octave-band noise exposure. These measurements were then repeated several weeks following noise damage, after the temporary effects of the exposure had stabilized. A preliminary review of this data was reported in an abstract for the ARO annual MidWinter Meeting (Martin *et al.*, 1998a).

II. METHODS

A. Experimental animals

A total of eight (16 ears) young adult pigmented rabbits were used in this study. Four rabbits were used to obtain pilot data on STC changes following a 2-kHz octave-band noise (OBN) exposure. Results from these animals indicated that in some cases, dramatic shifts in STC center frequencies (CF) could be observed. To explore this possibility further, STCs were systematically collected in four rabbits over a much larger frequency range. Results from the latter rabbits comprise the bulk of the data reported here.

Prior to any DPOAE testing, animals were administered an intramuscular injection of an anesthetic regimen (ketamine 50 mg/kg, xylazine 10 mg/kg), and surgically fitted with a permanent head-restraint device using aseptic methods. All rabbits were allowed at least 3 weeks to recover from this procedure before experimental manipulations were carried out. Throughout the course of the study, animals were housed within a standard vivarium facility, maintained on a 12:12-h light/dark cycle (lights on at 6:30 am), and provided with food and water *ad lib*. All experimental protocols were reviewed and approved by the Institutional Animal Care and Use Committee of the University of Miami School of Medicine.

B. DPOAE measurements

For all DPOAE measurements, awake rabbits were placed inside a standard plastic restrainer fitted with a bracket to which the surgically implanted headmount device was attached. Restrained rabbits were then placed inside a large, double-walled soundproofed-chamber. Each DPOAE measurement session typically included measurement of a set of DP-grams, i.e., DPOAE level as a function of frequency, with primary tones held at constant levels. Additionally, several interference response area (IRA) measurements were recorded, as described below.

The DPOAE measurement techniques employed in the present study have been described in detail previously (Whitehead *et al.*, 1995). Briefly, f_1 and f_2 primary tones were produced by two digital-to-analog channels of a 16-bit digital signal processing (DSP) board (Digidesign, Audiomedia), mounted in a personal microcomputer (Macintosh IIci). These primary-tone signals were transduced and presented in the closed ear canal by individual ER-2 (Etymotic Research) speakers. Sound-pressure levels in the ear canal were measured using an ER-10 (Etymotic Research) microphone system. The ear-canal sound was sampled and averaged ($n=4$) by an analog-to-digital channel of the DSP board. A 4096-point fast Fourier transform (FFT) of the time sample was then performed by customized software. DPOAE and noise-floor (NF) levels were extracted from the FFT. DP-grams were obtained at geometric-mean (GM) frequencies [$\mathrm{GM}=(f_1^* f_2)^{0.5}$], in 0.1-octave steps, from 1.414 to 18.37 kHz ($f_2=1.581-20.549$ kHz), with a fixed f_2/f_1 ratio of 1.25 for primary-tone levels below 65 dB SPL. For primary-tone levels $\geqslant$65 dB SPL, DP-grams had a high-frequency cutoff of 11.315 kHz, due to the dropoff of the maximal SPLs obtainable from the ER-2 speakers at high frequencies.

C. Interference response areas (IRAs)

IRAs were obtained for each subject using primary tones increasing in half-octave steps. Primary-tone GM frequencies were 1.414 ($f_1=1.27$, $f_2=1.59$ kHz), 2 ($f_1=1.79$, $f_2=2.24$ kHz), 2.828 ($f_1=2.53$, $f_2=3.16$ kHz), 4 ($f_1=3.59$, $f_2=4.47$ kHz), 5.6 ($f_1=5.06$, $f_2=6.33$ kHz), 8 ($f_1=7.15$, $f_2=8.94$ kHz), and 11.3 kHz ($f_1=10.12$, $f_2=12.66$ kHz). IRAs were collected using both equal and unequal primary-tone levels ($L_1=L_2=45$, 50, 55, 60, 65, 70, and 75 dB SPL; $L_1/L_2=50/35$, 55/45, and 60/55 dB SPL). These latter primary-tone combinations have been shown to be most sensitive to cochlear insult in rabbits (Whitehead *et al.*, 1995). As with the DP-gram protocol, primary-tone frequencies were held at a fixed ratio of $f_2/f_1=1.25$.

In order to alter the DPOAE output of the cochlea, an f_3 IT was digitally added to the f_1 channel of the DSP board. This tone was then increased systematically through 11 levels, from 35–85 dB SPL, in 5-dB steps. The IT was presented at 27 frequencies in 0.2-octave steps (i.e., covering a range of just over five octaves). The f_3 range was adjusted for different primary-tone frequency combinations such that suppression could be monitored with f_3 well over an octave below and above f_2 for most IRAs. IT frequency ranges for each set of primary-tone GMs were as follows: GM $=1.414$, $f_3=0.129-4.59$ kHz; GM$=2.0$, $f_3=0.172-6.07$ kHz; GM$=2.8$, $f_3=0.258-9.20$ kHz; GM$=4.0$, $f_3=0.323-12.1$ kHz; GM$=5.6$, 8, and 11.3, $f_3=0.495-18.4$ kHz.

Ear-canal sound-pressure levels for IRAs were sampled using the equipment and customized software described above. A 2048-point FFT was performed on the time sample.

While the phases of the primary tones were fixed relative to each other, the IT was presented at four different phase angles in 90° steps and subsequently time averaged. This protocol was used to reduce common first-order aliasing problems with emissions generated by the interaction of the IT with the primary tones.

Presentation of the IT at 27 frequencies and 11 levels produced a data matrix of 297 frequency/level combinations. At the end of each frequency column of the interference matrix, a control DPOAE was recorded in the absence of the IT. Thus, 27 control DPOAEs were recorded with each IRA matrix, which were then averaged in level to produce a reference control DPOAE level. The amount of interference was calculated by subtracting suppressed DPOAE levels from this reference level. When comparisons were made between IRAs generated by using control DPOAEs from each frequency column and the reference DPOAE level, IRAs were shown to be unaffected by this mean-reference procedure. This method was also useful in reducing noise in IRAs at times following noise exposure when reduced DPOAE levels approached the NF.

IRA contour plots were produced from each data matrix in successive 3-dB increments of DPOAE change (e.g., from 1.5–4.5 dB of suppression or enhancement). In the case of suppression, contours of increasing DPOAE decrement were calculated until the suppressed DPOAE level reached the NF.

D. Suppression tuning curves (STCs)

STCs were extracted from IRA contour plots as previously described in detail (Martin *et al.*, 1998b). Toward this end, a spreadsheet algorithm was used to create STCs at four suppression criteria of 3, 6, 9, and 12 dB. While STCs have a regular and predictable shape, enhancement response areas do not, and thus this algorithm was not used to calculate enhanced STCs. Tuning-curve tip CF, tip threshold, and the $Q_{10\,dB}$ measure of tip sharpness were automatically calculated for each of STC. Also extracted from these data sets were tip-to-tail differences [abbreviated as T_E by Mills (1998)]. To calculate this parameter, the IT required to induce 3- or 6-dB of DPOAE suppression was extracted from the data matrix both at the GM frequency and one octave below it. Values for T_E were then calculated as difference scores $[L_3(f_3=\text{GM-1 octave})\text{-}L_3(f_3=\text{GM})]$ between IT levels at these two frequencies, for each test run. In order to make a fair comparison between pre- and postnoise exposure T_E values, the primary-tone GM was selected as the standard tip frequency. This maneuver controlled for the small increases and decreases in CF seen in individual STCs following noise exposure. For all STCs, this frequency was still in the region of the tip, such that using the primary-tone frequency provided a fairly accurate measure of T_E. As several of the IRAs illustrated below show, the low-frequency tail often did now display a flat slope at one octave below the primary-tone GM. However, this GM-1 octave frequency was selected for calculation of T_E's in order to better compare the present data with previously published studies, and due to the fact that at frequencies lower than this, the 3-dB suppression criterion was not reached by levels within the IT stimulus matrix.

Preliminary analysis of DP-grams revealed that noise-induced decreases in DPOAE levels often made it impossible to obtain STCs at the higher suppression criteria. Thus, most of the comparisons made in the present study focused on the 3- and 6-dB STCs.

E. Experimental protocol

A preliminary set of DPOAE measurements consisting of DP-grams and IRAs was recorded from each subject prior to noise exposure. Rabbits were then placed in a large wire cage within a soundproofed booth. Each subject was exposed to a 105-dB SPL OBN, centered at 2 kHz, for 6 h. Rabbits were returned to their homecages overnight, and then exposed the next day to the same noise for another 6 h. In order to determine the effects of permanent noise-induced insults rather than temporary damage, all subjects were allowed to recover for a full 3 weeks following this exposure. An identical second set of DPOAE measurements was collected after the 3-week recovery period. Additional data sets were obtained from some subjects at 8 and 18 weeks, and 7 months following noise exposure in order to determine if any long-term recovery effects could be identified.

While no independent measure of noise-induced cochlear damage (e.g., auditory-evoked potentials, histological examination) was carried out in the present study, noise exposures similar to this have been shown to cause permanent cochlear dysfunction. For example, a recent study in chinchillas compared shifts in DPOAE level and auditory brainstem response with histopathological changes following an OBN exposure slightly higher in level but shorter in duration (Harding *et al.*, 2002). Other studies have made similar comparisons between DPOAE level, auditory-evoked potential threshold, and hair cell loss following narrow-band noise exposure in chinchilla (e.g., Hamernik *et al.*, 1998). While the noise-induced changes occurring in the cochlea cannot be precisely defined by examining only changes in DPOAE level, it is generally accepted that DPOAE level changes correlate with changes in cochlear function.

F. Data analysis

Values for the STC parameters of CF, tip threshold, and $Q_{10\,dB}$ for both pre- and postexposure measures from the second group of four rabbits were entered into a database. Values for each pre- and postexposure STC parameter were compared and analyzed using nonparametric paired sign tests (STATVIEW v.4.5). This statistical test was chosen because no assumptions were made about the underlying distributions of the data in the samples analyzed. While these tests become highly conservative with low numbers of data points, most data sets had an adequate number of measures for this test. Inspection of the data revealed considerable variability in control measures, suggesting that statistical analyses based upon treatment means would be unduly influenced by extreme scores. Thus, the more conservative sign test that considers only the direction of postnoise exposure change for each STC parameter was employed. This test assumes that all pre-exposure measures are equally likely to either increase or decrease after noise, if there is no treatment effect.

All IRAs recorded were examined for the presence of high-frequency (HF) interference lobes. The phenomenon of suppression and/or enhancement of DPOAE levels by f_3 tones with frequencies greater than f_2 has been described in detail previously (Martin *et al.*, 1999). For the purpose of objectivity, HF interference was considered present only when the control DPOAE level was changed (in either a positive or negative direction) by a 3-dB criterion, and these lobes were separated from the steep high-frequency slope of the main suppression lobe (i.e., at f_3 frequencies greater than 1/2 octave above f_2). It was also noted if the HF lobes of suppression or enhancement were affected by noise exposure, defined by increases in the f_3 level required to elicit HF lobes and by decreases in the amount of interference induced by f_3 (i.e., decreases in the amount of suppression or enhancement). When these lobes merged with the main lobe of suppression close to the f_2 frequency, a cutoff was manually set such that only contours with at least 10 dB of separation from the HF lobe would be scored.

III. RESULTS

A. Noise-induced DPOAE changes

Decreases in DPOAE levels were seen across a large frequency range in all rabbits following the OBN exposure. These decrements reflected permanent changes in the cochlea as DPOAE measurements recorded up to 9 weeks after noise exposure showed no further recovery than the data presented here. The most consistently damaged frequency region ranged from ~2.8–5.6 kHz. This frequency extent was equal to a range stretching from 1/2 to 1 1/2 octaves above the center of the OBN. When $L_1 = L_2 = 55$ dB SPL, the average DPOAE loss at the center of this frequency range, 4 kHz, was −12 dB (range = −9 to −35). This frequency pattern is consistent with previously published reports of noise-induced cochlear dysfunction (e.g., Slepecky, 1986). Each rabbit also exhibited DPOAE decrements in both ears at high frequencies above 6.5 kHz.

Figure 1 illustrates the differences between pre- and postexposure DP-grams (post minus pre) from each ear tested, using primary-tone levels of $L1 = L2 = 55$- (A), 65- (B), and 75-dB SPL (C). The postexposure measures shown here were obtained 4–5 weeks following noise exposure, and thus the differences illustrated can be regarded as permanent changes in DPOAE levels. Considerable variability existed in both the amount of DPOAE decrement induced the OBN, and in the frequency range across which DPOAEs were affected. This variability was observed both between rabbits and between the ears of the same rabbit. Interestingly, a second region of noise-induced DPOAE reduction was often observed at the higher test frequencies above 7–8 kHz. An additional feature of these measures was the presence of an enhancement of DPOAE level in some ears over the middle-frequency range, just above the region of consistent DPOAE loss. However, enhancement was found only when using high-level primary-tones to stimulate DPOAEs [Fig. 1(C)], and was most prominent in rabbit R45. Enhancements such as this have been described previously in rabbits following noise exposure (Franklin *et al.*, 1991; Howard *et al.*, 2002).

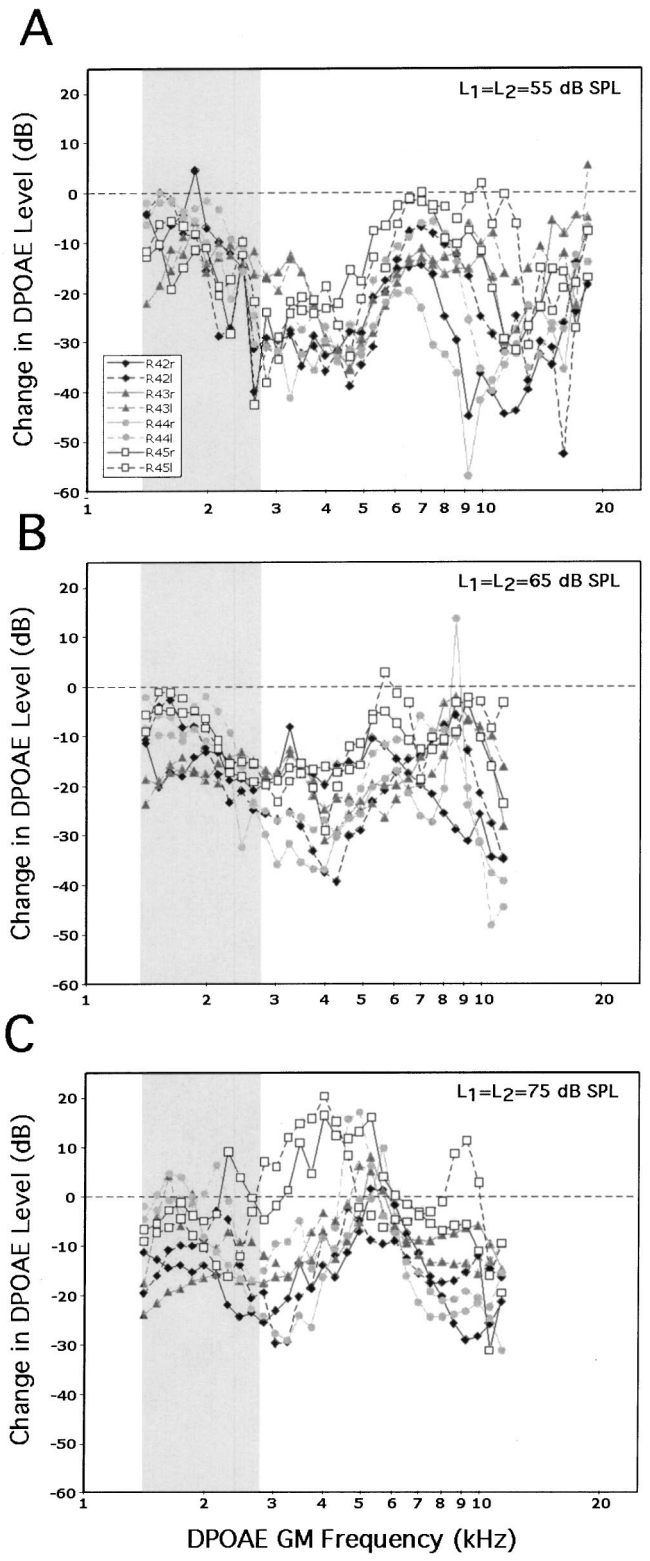


FIG. 1. DPOAE loss for each rabbit following noise exposure as a function of frequency. Data are displayed as ''difference'' DP-grams generated using primary tone levels of $L_1 = L_2 = 55$ (A); 65 (B); and 75 dB SPL (C). Difference scores were produced by subtracting pre- from postexposure DPOAE levels at each frequency in the DP-gram. While a degree of variability existed in both the level of DPOAE loss and the frequency extent over which the loss occurred, the region of greatest loss was generally in the range from 1/2 to 1 1/2 octave above the center of the OBN (stippled bar). Note a secondary region of notable DPOAE loss above ~6 kHz. DPOAE level enhancements can be seen in the data recorded using the highest primary-tone levels (C). The legend at the lower left of (A) identifies individual rabbits for all three plots. Primary-tone levels used for measuring each data set are listed in the upper-right corner of each plot.

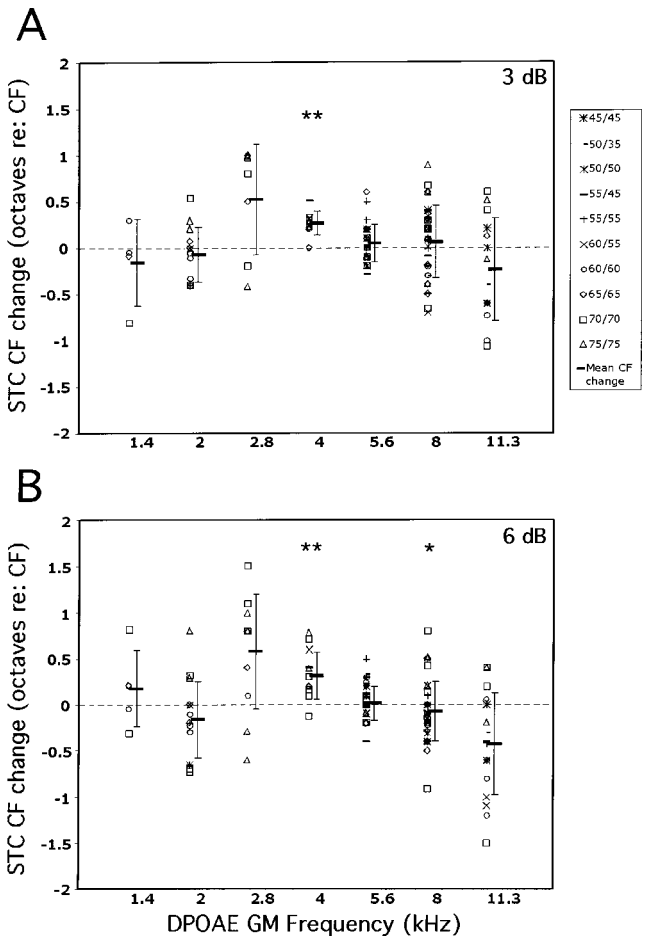


FIG. 2. Scatterplots of data for all comparisons, as well as the mean value, for changes in STC CF with respect to octaves above or below CF at each test frequency. Displayed are data from 3- (A) and 6-dB STCs (B). Pre-exposure values were subtracted from postexposure CFs. Different symbols indicate the primary-tone levels employed for individual comparisons, indicated in the legend at the right of (A). Bold horizontal dashes and error bars depict mean value ±1 s.d. Asterisks at the top of each plot indicate frequencies at which the average change between post- and pre-exposure values was significant ($*p<0.05$; $**p<0.01$).

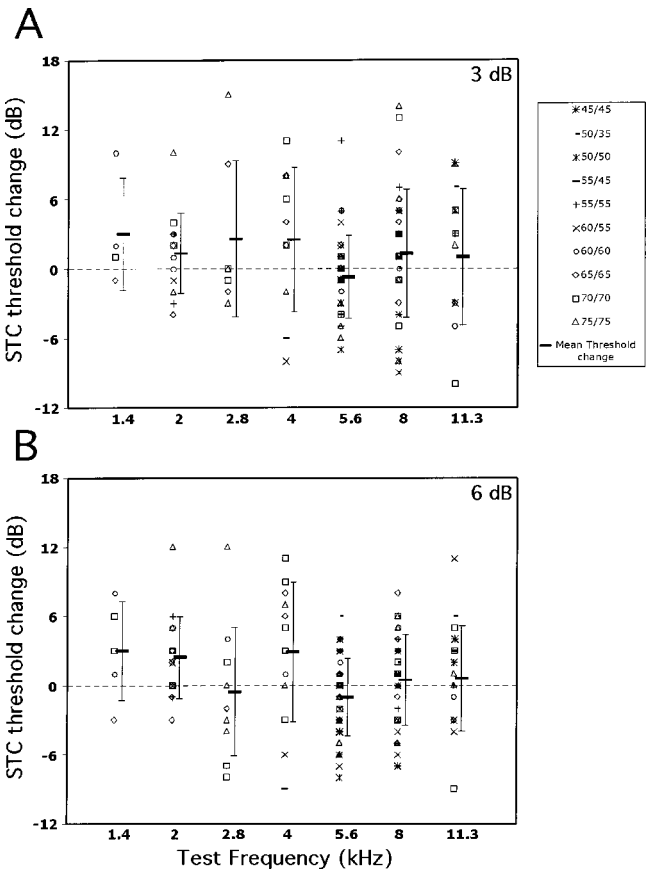


FIG. 3. Scatterplots and mean values of STC threshold change in dB at each test frequency following noise exposure. Individual data points represent differences post- minus pre-exposure values for STC threshold for the 3- (A) and 6-dB (B) functions. Details and the legend for these plots are as in Fig. 2.

B. Noise-induced STC changes

Because DPOAE levels were substantially altered in each rabbit, responses above the NF could be elicited at most frequencies only by using higher primary-tone levels. Thus, most of the pre- versus postexposure comparisons were restricted to STCs collected at the higher levels. Additionally, more data were collected at frequencies well below or above the 4-kHz frequency region that displayed the largest DPOAE decrements.

Figure 2 illustrates changes in CF following noise exposure for 3- (A) and 6-dB (B) STCs as a function of primary-tone GM frequency. These plots show both individual measures of STC parameters, indicated by symbols matching the primary-tone levels used for each data point, and the mean post- minus pre-exposure value for all data points [error bars indicate ±1 standard deviation (s.d.)]. It can be seen that, overall, a great deal of variability existed in STC behavior following noise exposure. In the frequency regions involving the greatest DPOAE decrements, between ~2.8–4 kHz (see Fig. 1), both 3- and 6-dB STCs showed a tendency to tune to higher frequencies. Statistical analysis of these data revealed a significant effect, however, only at 4 kHz in each of these two data sets for tuning shifts to be towards positive frequencies, i.e., for the CF to shift to higher frequencies. In contrast, statistical tests on the 6-dB STCs obtained using a primary-tone GM of 8 kHz revealed a significant tendency for these STCs to tune to lower frequencies following noise exposure. While some STCs recorded with other primary-tone GM frequencies tuned to higher or lower CFs following noise exposure, these changes were not significant based upon the direction of change (i.e., as analyzed by the non-parametric sign test).

Figure 3 displays the STC threshold changes for each function examined, as well as the mean post minus pre change, as a function of primary-tone GM frequency for 3- (A) and 6-dB (B) STCs. As with changes in the CF, a great degree of variability existed in how noise exposure affected this DPOAE parameter, with STC thresholds shifting equally to higher and lower levels at most GM frequencies. Consequently, none of the noise-induced modifications in this parameter was statistically significant.

Figure 4 illustrates changes in the $Q_{10\,dB}$ measure of tuning curve sharpness induced by acoustic overstimulation. As with Figs. 2 and 3, differences between post- and pre-exposure measurements are displayed for 3- (A) and 6-dB (B) STCs. There was a large degree of variability in how this

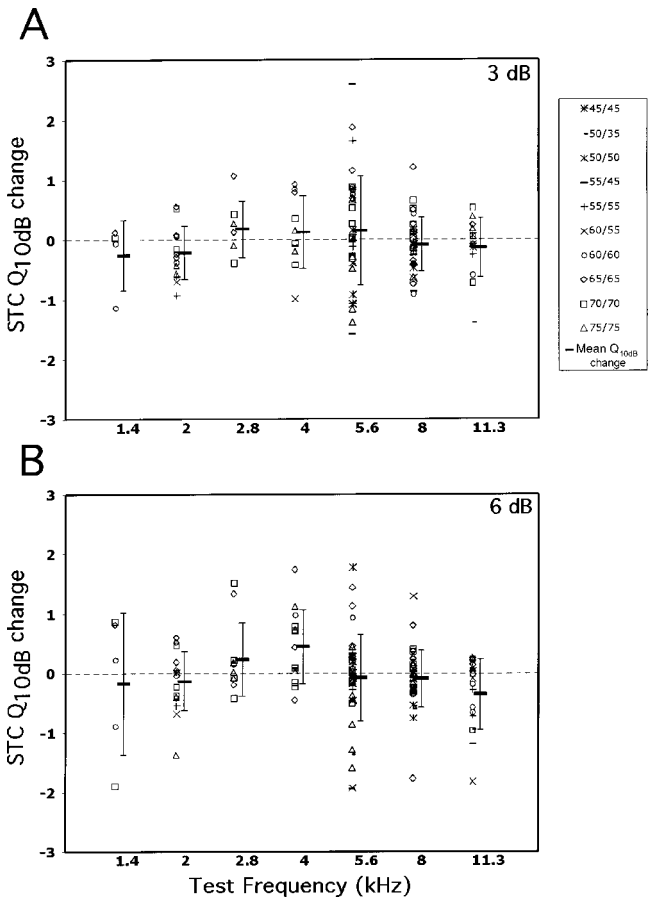


FIG. 4. Scatterplots and mean values of STC $Q_{10\,dB}$ change at each test frequency following noise exposure. Individual data points represent differences post- minus pre-exposure values for STC $Q_{10\,dB}$ for the 3- (A) and 6-dB (B) functions. Details and the legend for these plots are as in Fig. 2.

STC parameter was affected following OBN exposure. Both positive and negative shifts in $Q_{10\,dB}$ can be seen in the individual data points associated with each test frequency. Statistical analyses of these data again revealed no significant changes between pre- and postexposure $Q_{10\,dB}$ values.

Figure 5 illustrates two examples, taken from two different rabbits, of the typical postexposure upward CF shifts at 4 kHz described above. The top four panels [Figs. 5(A)–(D)] display data taken from the right ear of subject R42, at the 4-kHz test frequency, using primary-tone levels of $L_1 = L_2 = 65$ dB SPL. Panels on the left side display the IRAs recorded before (A) and 3 weeks after (C) OBN exposure. Next to each of these (B), (D) are the counterpart STCs extracted from the data matrices. Control DPOAE levels are displayed at the top right of each STC. Inset at the left bottom of each STC are the CF values extracted for 3- and 6-dB STCs. Arrows on the abscissa of each plot represent the primary-tone levels employed in collecting the data. The second example shown in Figs. 5(E)–(H) was taken from the right ear of subject R44 at the same test frequency, using primary tone levels of $L_1 = L_2 = 70$ dB SPL. Pre- (E) and postexposure IRAs (G), and their accompanying STCs (F) and (H) are arranged as in the first example. Comparison of these two sets of STCs reveals an upward shift of CF in the presence of large (~20 dB) permanent noise-induced DPOAE decrements in both subjects. Although these CF shifts are not as dramatic as those seen in the four pilot animals (see below), a clear change in the trajectory of all the contour lines can be seen between the pre- versus postexposure IRAs.

Representative data from test frequencies lower than 4 kHz are illustrated in Fig. 6. The left column of these plots illustrates pre-exposure IRAs, while the corresponding postexposure measures are shown in the right column. Data are displayed for the test frequencies at 1.4 (A), (B); 2 (C), (D); and 2.8 kHz (E), (F). The unique subjects from which each data set was recorded is listed above (left) each IRA and STC. Small changes in STC parameters occurred when comparisons between these individual examples were made, but, as described above, these changes were not systematic across the population tested.

Figure 7 illustrates data from test frequencies higher than 4 kHz. This figure is arranged in the same way as Fig. 6 for the 5.6- (A), (B); 8- (C), (D); and 11.3-kHz (E), (F) test frequencies. As with Fig. 6, small changes can be seen between pre- and postexposure IRAs, due to the natural variability of noise-induced cochlear dysfunction, but these modifications were not systematic across the data obtained at these test frequencies.

C. Noise-induced changes in HF lobes

A total of 158 pairs of pre- and postexposure IRAs was examined for the presence of HF interference lobes. The phenomenon of HF interference occurred in some rabbits at almost all primary-tone frequency and level combinations, but was elicited most commonly when higher-level primary tones were used (e.g., $L_1 = L_2 = 60$ dB SPL and above). Of the pre-exposure IRAs, 101 (64%) exhibited HF suppression lobes, and 46 (29%) exhibited HF enhancement lobes. Nine IRAs showed both HF enhancement and suppression lobes. Comparisons made between pre- and postexposure IRAs showed both HF suppression and enhancement lobes to be highly sensitive to noise exposure. Both the size and the depth (i.e., the frequency extent of the area of interference and the magnitude of interference, respectively) of interference were greatly reduced in 53% of HF suppression lobes and in 76% of HF enhancement lobes.

Noise-induced changes in HF interference lobes can be seen in Figs. 5–7. For example, Fig. 5(A) shows a small area of suppression centered approximately one octave above f_2 at ~9 kHz that was >3 dB. Following noise exposure, this HF suppression lobe was no longer present, as shown in Fig. 5(C).

Other examples of this phenomenon can be seen when comparing pre- and postexposure IRAs at the 1.4- and 2-kHz test frequencies [Figs. 6(A)–(D)]. These data sets reveal high degrees of HF enhancement approximately 1/2 octave above f_2 and large HF suppression lobes approximately one octave above f_2. Following noise exposure, these HF interference features were virtually nonexistent. At the higher frequencies displayed in Fig. 7, HF suppression lobes were diminished in size following noise exposure at the test frequencies of 5.6- (A) vs (B) and 8 kHz (C) vs (D). While large changes in HF lobes are not shown in the 2.8- or 11.3-kHz examples, such changes did occur at these frequencies as well.

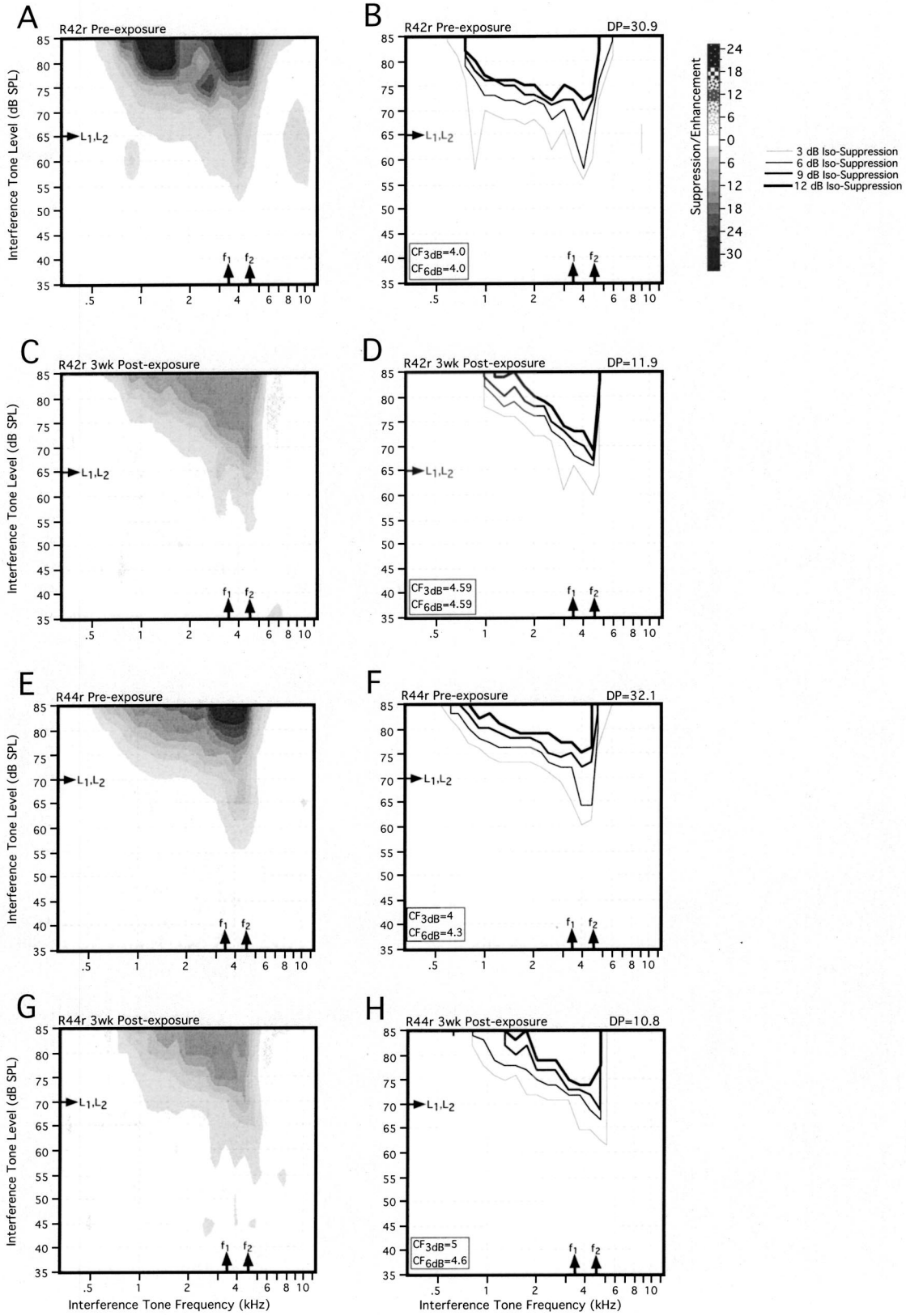


FIG. 5. Postexposure changes in STC CFs at the 4-kHz test frequency. IRAs (left panels) and STCs (right panels) from two subjects are displayed. Pre-exposure measures for the right ear of subject R42 (A), (B) are compared to their corresponding postexposure measures (C), (D) recorded using primary-tone levels of $L_1 = L_2 = 65$ dB SPL. Pre- (E), (F) and postexposure (G), (H) data for the right ear of subject R44 were recorded using primary-tone levels of $L_1 = L_2 = 70$ dB SPL. In each of these examples, CF increased in the presence of large ($\geqslant$19 dB) permanent noise-induced DPOAE reductions. Control DPOAE levels are listed at the top right of each STC panel. Values for the CF parameter of both 3- and 6-dB STCs are inset at the lower left of each STC plot. Primary-tone levels are indicated by the arrows on the ordinate and frequencies by arrows on the abscissa.

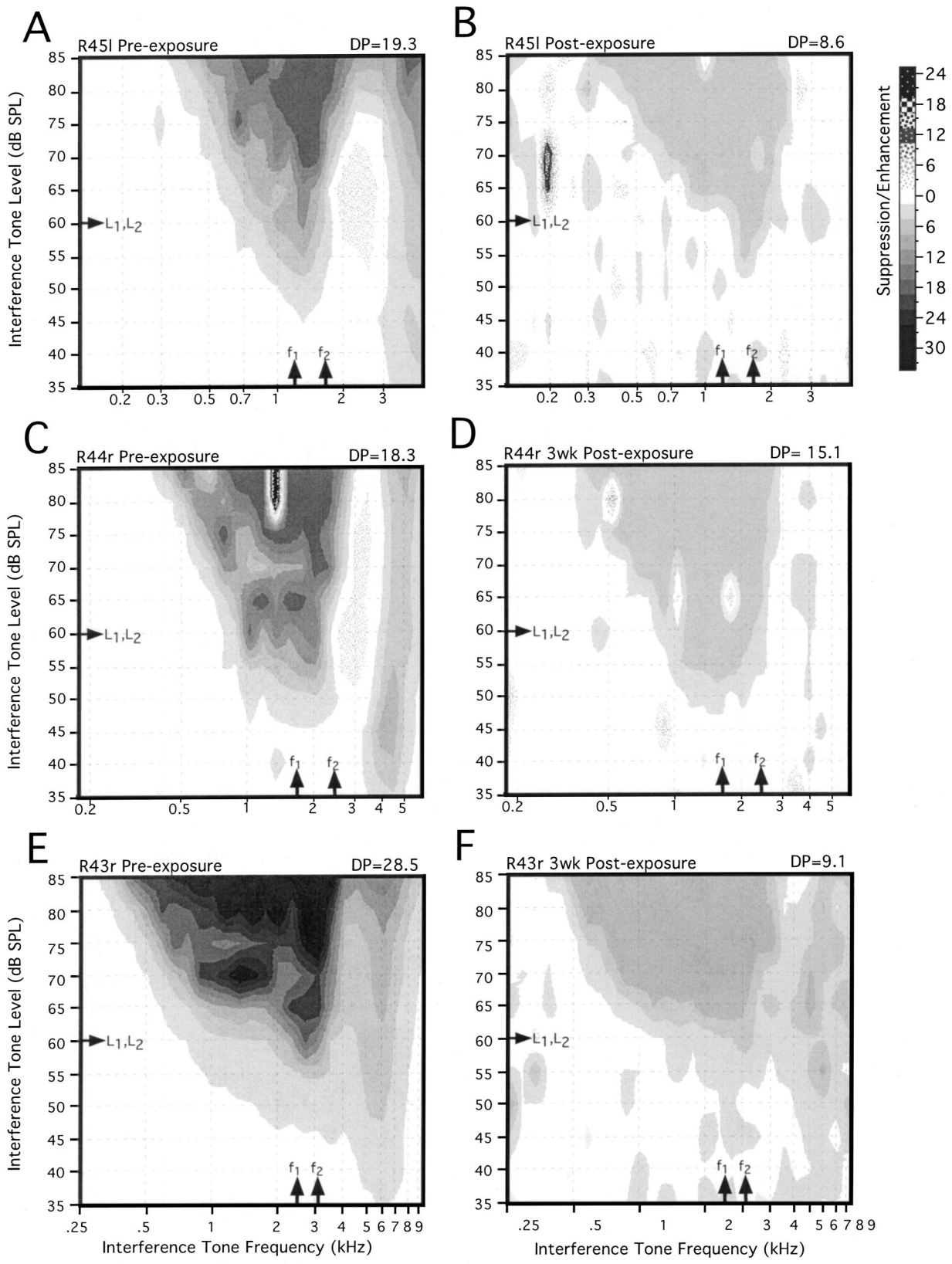


FIG. 6. Representative pre- (left panels) and postexposure IRAs (right panels) recorded at test frequencies lower than 4 kHz. Data are shown for test frequencies at 1.4 (top row), 2 (middle row), and 2.8 kHz (bottom row). These examples were taken from three different subjects (R45 left ear, R44 right ear, R43 right ear, respectively), using primary-tone levels of $L_1 = L_2 = 60$ dB SPL. Comparison between pre- and postexposure plots reveals large changes in HF enhancement and suppression areas at the 1.4- and 2-kHz test frequencies. While the depth of suppression in each of these examples was affected, the slight changes between pre- and postexposure STC parameters were not systematic across all subjects (specific data not shown).

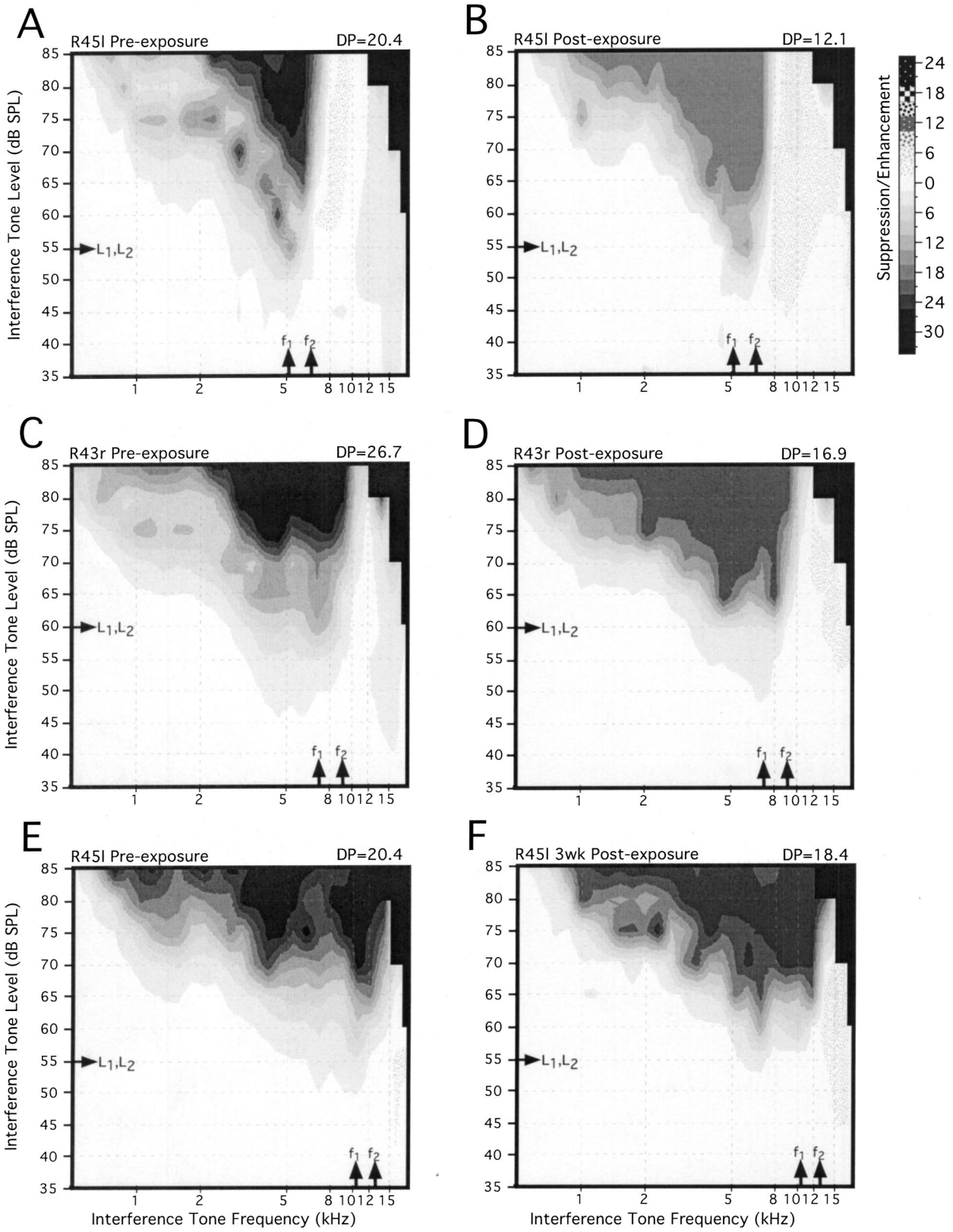


FIG. 7. Representative pre- (left panels) and postexposure IRAs (right panels) recorded at test frequencies higher than 4 kHz. Data shown are for test frequencies at 5.6 (top row), 8 (middle row), and 11.3 kHz (bottom row). These examples were selected from two different subjects [R45 left ear in (A), (B), (E) and (F); R43 right ear in (C) and (D)], using primary-tone levels of $L_1 = L_2 = 55$- and $L_1 = L_2 = 60$-dB SPL. Comparison between pre- and postexposure plots reveals large changes in HF enhancement and suppression areas at the 5.6- (A), (B) and 8-kHz (C), (D) test frequencies. As in Fig. 6, none of the slight changes in STC parameters recorded at these frequencies was systematic, despite permanent noise-induced DPOAE decrements and decreases in the depth of suppression. Blackened regions in areas of high IT frequency and level indicate where data could not be collected due to the dropoff of the speaker level at the higher frequencies.

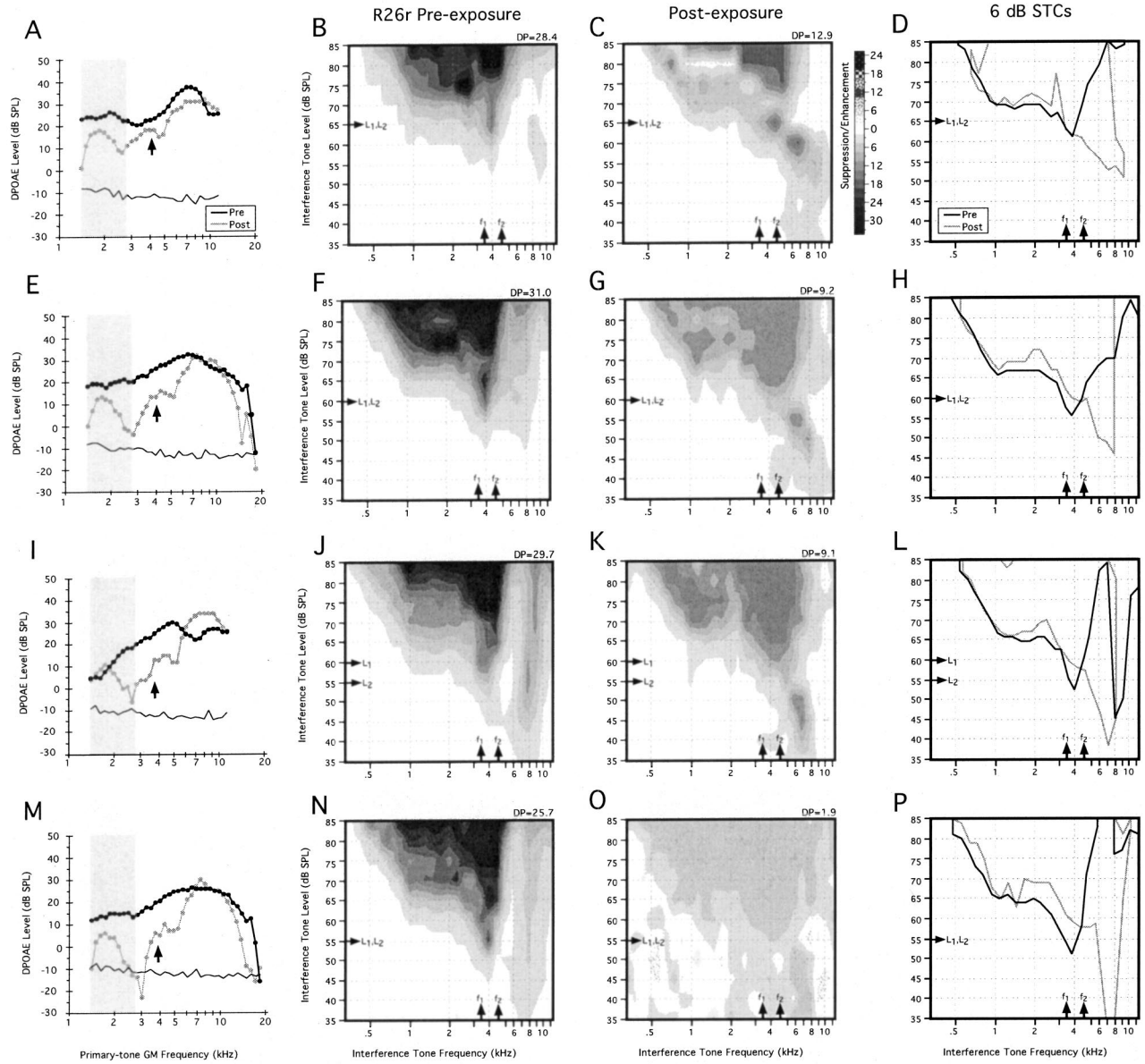


FIG. 8. Effects of primary-tone level on abnormal STC tuning for the right ear of R26 obtained at a GM of 4 kHz. The column on the left depicts pre- and post-DP-grams, obtained using primary-tone levels identical to the IRAs to the right [i.e., $L_1=L_2=65$ (top), 60 (second row), $L_1=60$, $L_2=55$ (third row), and $L_1=L_2=55$ dB SPL (bottom row)]. In this ear, there was a remaining region of clearly normal DPOAEs revealed by the $L_1/L_2=55/55$ DP-gram (lower left). The middle columns depict pre- and post-IRA contour plots, while the column on the far right shows pre- and post-STCs extracted for a suppression criterion of 6 dB. It is clear that, for all primary-tone levels, postexposure STCs tuned to more basal frequencies. At primary-tone levels of 65/65 dB SPL and lower, STC thresholds also became abnormal in that they were much lower than the eliciting primary tones. Below primary-tone levels of 55/55 dB SPL, no postexposure DPOAEs were measurable. Normal rabbits often show substantial suppression above f_2, as seen for this ear at $L_1/L_2=60/55$ dB SPL. It is interesting to note that this region of secondary suppression corresponded to the frequency region of the new, postexposure STC-tip frequencies.

Sometimes remnants of these HF lobes remained following noise exposure that may help explain the tendency for upward shifts in CF [e.g., Fig. 6(F)]. This finding was particularly true for several rabbits in the pilot group that showed very dramatic CF changes. Figure 8 illustrates this finding, which is particularly clear when IRAs are viewed as a primary-tone level series. The left column of this figure shows pre- (solid circles) and postexposure (gray circles) DP-grams in order of descending primary-tone levels. The arrow in each plot indicates the 4-kHz GM frequency used to collect the associated pre- and postexposure IRAs in the two middle columns and the corresponding 6-dB suppression criterion STCs in the right column. Inspection of these DP-grams reveals that the 4-kHz place lies in a region of substantial dysfunction, especially apparent in the DP-grams collected at the lower primary-tone levels. Interestingly, this damaged region is flanked on the high-frequency side (6–10 kHz) with a region of essentially normal DPOAEs, as indicated by the DP-grams in this region. The pre-exposure IRAs for the right ear of this rabbit (R26r) revealed a large-HF lobe, especially at primary-tone levels of 60/55 and 55/55 dB SPL. Postexposure, this HF lobe was still prominent except at the lowest (55/55) primary-tone level. The presence of this lobe appears to be responsible for the postexposure STCs tuning to higher frequencies (right column).

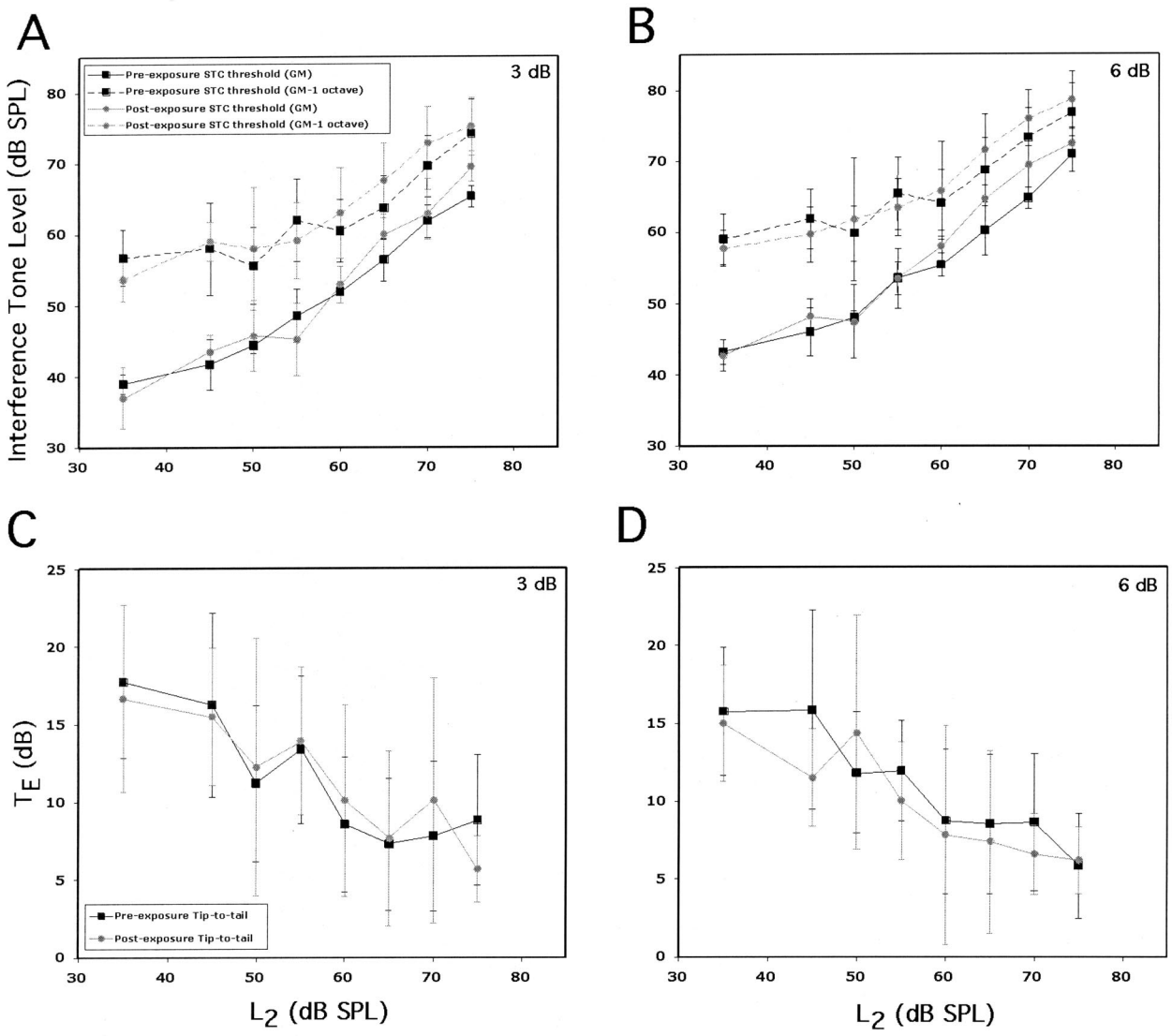


FIG. 9. IT levels required to reach criterion suppression levels at GM and GM minus one octave. Mean (±1 s.d.) IT levels required to reach suppression criteria of 3- (A) and 6-dB (B) as a function of L_2 are illustrated in the upper panels. The lower panels display the corresponding tip-to-tail difference, T_E, for each L_2 displayed. As primary-tone levels increased there was a decrease in T_E that would be expected as the gain of the CA became saturated at these higher stimulus levels. A change in T_E was not observed postexposure, because the gain of the CA operates on both f_2, where the DPOAE is produced, and f_3, when it is close to the f_2 place. Solid lines represent data from the primary-tone GM frequency, while dashed lines illustrate corresponding data for one octave below GM. Pre-exposure data are plotted using black symbols and lines; postexposure data are plotted using gray symbols and lines. The legend inset in (A) is also applicable to (B); the legend inset in (C) is also applicable to (D).

D. Noise-induced T_E changes

Further characteristics of third-tone interference within the main lobe of suppression were examined by calculating and comparing T_E's. Only primary-tone frequency/level combinations at which STCs could be recorded both before and after noise were used for these analyses. These data were averaged across test frequencies, and mean values (±1 s.d.) were plotted as a function of L_2 in Fig. 9 for the 3- (A) and 6-dB (B) suppression criteria. Changes in T_E as a function of L_2, corresponding to each suppression criterion level, are also shown in Figs. 9(C) and (D). While a degree of variability exists in these data, clear trends are illustrated.

In both pre- (black squares) and postexposure data (gray squares), the IT level required to meet the suppression criterion increased more rapidly at the GM test frequency, that is, at the tip region, than in the lower frequency (GM-1 octave) tail region. This trend led to decreases in the T_E with increases in primary-tone level [Figs. 9(C), (D)]. IT levels required to meet the suppression criteria at the GM and GM-1 octave frequencies, however, did not change following noise exposure. Thus, comparison between T_E's in STCs recorded before and after noise exposure showed no differences in magnitude or level-dependent behavior. When separated and analyzed by frequency, these effects were not altered, even at the 4- and 5.6-kHz test frequency that showed significant shifts in tuning (data not shown).

IV. DISCUSSION

The purpose of this study was to provide both qualitative and quantitative descriptions of the effects of permanent noise-induced injury on DPOAE suppression tuning. While a degree of variability existed between the extent of damage

each rabbit suffered following noise exposure, each subject exhibited long-lasting changes in DPOAE levels over similar frequency ranges. Similarly, variability existed in how DPOAE suppression measurements were altered following noise exposure. As with reversible cochlear insults such as administration of loop diuretics (Martin *et al.*, 1998b) and exposure to noise causing temporary DPOAE changes (Howard *et al.*, 2002), this acoustic overexposure did not disrupt DPOAE suppression tuning in the same way that neural or BM tuning would presumably be affected by a similar insult. This is reflected by the fact that neither the STC threshold, nor the $Q_{10\,dB}$ parameter, exhibited significant changes following noise exposure at any test frequency.

However, the CF of STCs recorded at 4 kHz was significantly affected following noise exposure in the rabbits tested. This basal shift in tuning was highly significant for both the 3- and 6-dB STC criteria. This test frequency corresponds with the region of greatest cochlear damage following noise exposure. Additionally, at 8 kHz, there was significant tendency for 6-dB STCs to tune to lower frequencies. These changes were not accompanied by systematic changes in any other STC parameter.

Just as STC tip thresholds did not change following noise-induced cochlear damage, the IT levels required to reach suppression criteria at the GM frequency, or below, also did not change. This is apparent in Fig. 9, in which T_E's were unaltered following noise exposure. The IT levels necessary for induction of criterion amounts of DPOAE suppression at the GM frequency and one octave below in rabbits, upon which T_E's are based, were similar to other recent reports of this type of data obtained in humans (Gorga *et al.*, 2002; Pienkowski and Kunov, 2001). Slight differences may be due to differences in species, stimulus parameters, or the frequencies selected for these analyses, as well as the inherent variability in this type of measurement.

The STC T_E has been hypothesized to be an indirect measure of cochlear amplifier (CA) gain (Abdala, 1998, 2001; Mills, 1998; Pienkowski and Kunov, 2001; Gorga *et al.*, 2002) based upon properties of BM vibration at and below CF. Similar arguments have been proposed to explain the reason why L_2 needs to be lower than L_1 so as to equate amplitudes on the BM (see Fig. 4, Kummer *et al.*, 2000). Thus, DPOAE STCs gain their characteristic shape from the linear and nonlinear properties of BM vibration. Fahey *et al.* (2000) predicted that DPOAE suppression is dependent upon the ratio of stimulus-tone levels at the site of DPOAE generation (i.e., DPOAE suppression will begin to occur as L_3 approaches within 10 dB of the level of L_2). In order to suppress DPOAE generation, BM vibration at the generation site (f_2) must be driven by f_3. When f_3 is much lower than f_2, L_3 growth is linear at the f_2 place (i.e., not affected by the CA), similar to the BM response. Thus, at frequencies below CF, much higher levels of the IT are required to produce suppression. This features accounts for the ''tail'' of the STC. As f_3 approaches f_2, both tones are subject to the same amplification, and the gain of each is subject to the compressive nonlinearity of the CA. This nonlinear growth of L_3 at the site of DPOAE generation, i.e., at the f_2 place, may account for the much lower threshold of the tip region of the STC, just as this nonlinearity accounts for the frequency selectivity shown by BM and NTCs. Because, the CA is most active at lower levels, increasing total energy (i.e., increasing L_2) saturates the CA, leading to smaller T_E's at higher primary-tone levels.

The data presented here, as well as in prior reports of previous experiments from this series of studies (Martin *et al.*, 1998b; Howard *et al.*, 2002), is in agreement with the above hypothesis in that gain (T_E) decreases as L_2 levels are increased. However, in damaged ears, the large differences in STC tip threshold that would lead to changes in this parameter were not present in this study or in the two previous studies. In the case of permanently damaging noise exposure, which has been shown to disrupt the sensitive tip of BM and neural tuning (i.e., a direct measure of the activity of the CA), neither changes in T_E's nor in primary-tone level-dependent behavior were observed. Thus, while this type of insult, as well as the manipulations employed by the previous studies described above (furosemide and temporarily damaging noise exposure), disrupts a ''direct'' measurement of CA gain, the current STC paradigm was ineffective at detecting changes in CA gain as estimated by T_E in damaged ears.

This inability of STCs to detect changes in the gain of the CA following cochlear insults can be explained by the fact that, as the IT approaches CF, the gain of the CA is approximately equal for both f_3 and f_2. Following cochlear damage, when the gain of the CA is reduced, this reduction in gain will affect both stimuli. Thus, even in a damaged cochlea, when f_3 is close to the frequency of the DPOAE generation site, DPOAE suppression will occur when L_3 comes within 10 dB of L_2, and the T_E will not change. This hypothesis is supported by the lack of change in T_E's in this study, and the lack of large increases in STC tip threshold presented here, as well as following both ototoxic poisoning (Martin *et al.*, 1998b) and temporarily damaging noise exposure (Howard *et al.*, 2002). Together, these results are in contrast to the recent report of a significant correlation between T_E's and hearing thresholds in humans (Pienkowski and Kunov, 2001). However, the results of the latter study may be confounded by the strong correlation of T_E with DPOAE level.

The HF interference lobes exhibited in the IRAs recorded in this study show similar characteristics to those previously described (Martin *et al.*, 1999; Fahey *et al.*, 2000; Mills, 2000). Most commonly, lobes of suppression occurred when f_3 reached frequencies about an octave above f_2. HF enhancement lobes often occurred at frequencies just above the main lobe of suppression. Some cases exhibited separate main and HF suppression lobes separated by a region of enhancement. This phenomenon was prevalent in IRAs at all test frequencies. As in earlier work, these regions of interference were highly susceptible to noise exposure, often changing dramatically or disappearing completely after the insult. These noise-induced changes in HF interference occurred as often in IRAs at test frequencies above the main region of damage as they did for IRAs recorded at test frequencies at, or below, the most damaged regions of the cochlea.

It has been proposed that HF lobes result from even-

order nonlinearities in which harmonics of the primary tones produce difference-tone emissions that coincide with the DPOAE being measured (Fahey *et al.*, 2000). This phenomenon may explain the upward change in CF observed in some rabbits. For example, if primary-tone harmonics are responsible for the HF lobes, then depending upon the pattern of damage, harmonics may be enhanced by changes in cochlear tuning. In these instances (see Fig. 8), much of the measured emission may come from the harmonic place and the apparent CF of the STC will be shifted to a higher frequency. Overall, the amount and pattern of damage may determine the tendency for upward shifts in tuning. As mentioned earlier, these lobes are typically very sensitive to damage, suggesting that the generation of even-order emissions may provide a sensitive indicator of cochlear abnormality in many circumstances.

The data presented illustrate the difficulty in using DPOAE suppression measurements to monitor cochlear pathology. Even under normal conditions, STC parameters in humans are influenced by primary-tone level (Kummer *et al.*, 1995). For example, many studies have shown that the $Q_{10\,dB}$ parameter decreases as L_2 increases (e.g., Harris *et al.*, 1992; Cianfrone *et al.*, 1994; Kummer *et al.*, 1995; Abdala *et al.*, 1996; Gorga *et al.*, 2002). While the data presented here are not analyzed in this way, T_E measurements clearly showed that suppression thresholds in the tip region and the tail region converge as L_2 increased, implying that the tuning curve is broadening (see Fig. 9). As with any cochlear insult that leads to decreases in DPOAE level, permanent noise-induced injury decreased the depth of suppression that can be measured before the NF is reached. Thus, it becomes increasingly difficult to collect data using low-level primary tones, which have been shown repeatedly to be more sensitive to cochlear insult than high-level DPOAEs (e.g., Whitehead *et al.*, 1995). An alternative paradigm would involve increasing the primary-tone levels until the pre-exposure DPOAE level is reached. With this particular procedure, it is predicted that changes in STC tuning would be similar to those observed for BM or NTCs.

While some individual STCs recorded exhibited obvious noise-induced changes following exposure (e.g., individual data points in Figs. 2–4), examination of group data shows that many of these effects are not systematic. This outcome was likely due in part to the variability seen in the amount and frequency range of noise-induced damage between subjects. However, the systematic changes in CF illustrated in the data presented in this study occurred in the region of greatest noise-induced cochlear change, indicating a reflection of noise-induced damage in an STC parameter. Abnormalities in STC parameters have been described under several conditions including in premature neonatal humans (Abdala, 1998; Abdala *et al.*, 2001), following temporarily damaging noise (Howard *et al.*, 2002), and now following permanently damaging noise. Similarly, STC abnormalities have also been described in human children with mild to moderate sensorineural hearing loss (Fitzgerald and Abdala, 2002). However, due to the small magnitude of most of the described changes, these STC parameters remain a difficult tool to use for diagnosing cochlear dysfunction in individuals. While individual examples of STC change following cochlear insult may illustrate a reflection in change in cochlear function, the variability and lack of systematic change in group data makes it difficult to understand the mechanisms leading to changes in these functions.

V. CONCLUSIONS

DPOAE suppression tuning characteristics have been described following a variety of different cochlear insults and pathologies. Insults that affect cochlear frequency tuning, as classically defined as disrupting the frequency selectivity of BM and/or neural tuning, often do not alter the properties of STCs in the same way. Data presented here also showed no significant changes in STC thresholds or $Q_{10\,dB}$ values following noise exposure. These data are in line with previously described rules for suppression of saturating nonlinearities (Fahey *et al.*, 2000) and can be explained because the gain for both f_3 at the f_2 place, and the gain for f_2 are changed simultaneously. Thus, the traditional STC technique cannot detect changes in the gain of the CA following cochlear damage when primary-tone levels are held constant. However, this technique appears to reflect the reduction in gain of the CA due to saturation when the levels of the primaries are increased. The present study revealed significant systematic changes in the CF of STCs recorded in the region of greatest noise-induced cochlear damage. Other types of STC abnormality have been reported previously in premature human neonates (Abdala, 1998) and during noise-induced temporary threshold shifts (Howard *et al.*, 2002). Each of these studies reported increases in $Q_{10\,dB}$ in the experimental condition. A large percentage of the pre-exposure IRAs examined also exhibited HF suppression and enhancement lobes. These areas of interference were highly susceptible to noise-induced dysfunction, mirroring results reported previously (Martin *et al.*, 1999). Due to the small size of the changes in STC parameters reported in this and previous studies, it is difficult to use these measures as a diagnostic tool for cochlear dysfunction. However, examining such functions in a population before and after controlled cochlear insults, such as noise exposure, may provide information about the specific characteristics of the cochlear nonlinearity.

ACKNOWLEDGMENTS

The authors wish to acknowledge the two anonymous reviewers whose comments helped to improve the quality of this manuscript. This research was supported in part by funds from the Public Health Service (DC00613, ES03500), and funds from the University of Miami Chandler Chair. M. A. Howard is a Lois Pope LIFE fellow.

Abdala, C. (**1998**). ''A developmental study of distortion product otoacoustic emission ($2f_1-f_2$) suppression in humans,'' Hear. Res. **121**, 125–138.

Abdala, C. (**2001**). ''Maturation of the human CA: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels,'' J. Acoust. Soc. Am. **110**, 1465–1476.

Abdala, C., Visser-Dumont, L., and Ma, E. (**2001**). ''DPOAE suppression tuning: Cochlear immaturity or aging effects in adults?'' Abstr. Assoc. Res. Otolaryngol. **24**, 10.

Abdala, C., Sininger, Y. S., Ekelid, M., and Zeng, F.-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the 2 f_1-f_2 stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Cianfrone, G., Altissimi, G., Cervellini, M., Musacchio, A., and Turchetta, R. (**1994**). ''Suppression tuning characteristics of 2 f_1-f_2 distortion product otoacoustic emissions,'' Br. J. Audiol. **28**, 205–212.

Evans, E. F., and Klinke, R. (**1982**). ''The effects of intracochlear and systemic furosemide on the properties of single cochlear nerve fibers in the cat,'' J. Physiol. (London) **331**, 409–428.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). ''Nonlinear interactions that could explain distortion product interference response areas,'' J. Acoust. Soc. Am. **108**, 1786–1802.

Fitzgerald, T. S., and Abdala, C. (**2002**). ''Ipsilateral DPOAE (2 f_1-f_2) suppression in children with hearing loss: Preliminary data,'' Abstr. Assoc. Res. Otolaryngol. **25**, 776.

Frank, G., and Kossl, M. (**1995**). ''The shape of 2 f_1-f_2 suppression tuning curves reflects basilar membrane specialization in the mustached bat, *Pteronotus parnellii*,'' Hear. Res. **83**, 151–160.

Franklin, D. J., Lonsbury-Martin, B. L., Stagner, B. B., and Martin, G. K. (**1991**). ''Altered susceptibility of 2 f_1-f_2 acoustic-distortion products to the effects of repeated noise exposure in rabbits,'' Hear. Res. **53**, 185–208.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Konrad-Martin, D. (**2002**). ''The use of distortion product otoacoustic emission suppression as an estimate of response growth,'' J. Acoust. Soc. Am. **111**, 271–284.

Hamernik, R. P., Ahroon, W. A., Jock, B. M., and Bennett, J. A. (**1998**). ''Noise-induced threshold shift dynamics measured with distortion-product otoacoustic emissions and auditory evoked potentials in chinchillas with inner hair cell deficient cochleas,'' Hear. Res. **118**, 73–82.

Harding, G. W., Bohne, B. A., and Ahmad, M. (**2002**). ''DPOAE level shifts and ABR threshold shifts compared to detailed analysis of histopathological damage from noise,'' Hear. Res. **174**, 158–171.

Harris, F. P., Probst, R., and Xu, L. (**1992**). ''Suppression of the 2 f_1-f_2 acoustic distortion product: changes with primary level,'' J. Acoust. Soc. Am. **64**, 133–141.

Howard, M. A., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**2002**). ''Effects of reversible noise exposure on the suppression tuning of rabbit distortion-product otoacoustic emissions,'' J. Acoust. Soc. Am. **111**, 285–296.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). ''Basilar membrane measurements and the travelling wave,'' Hear. Res. **22**, 147–153.

Kiang, N. Y. S., Liberman, M. C., and Levine, R. A. (**1976**). ''Auditory-nerve activity in cats exposed to ototoxic drugs and high-intensity sounds,'' Ann. Otol. **85**, 752–768.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA), Res. Monogr. 35.

Koppl, C., and Manley, G. A. (**1993**). ''Distortion-product otoacoustic emissions in the bobtail lizard. II. Suppression tuning characteristics,'' J. Acoust. Soc. Am. **93**, 2834–2844.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the 2 f_1-f_2 distortion-product emissions in humans,'' J. Acoust. Soc. Am. **98**, 197–210.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). ''Optimal L_1-L_2 primary tone level separation remains independent of test frequency in humans,'' Hear. Res. **146**, 47–56.

Liberman, M. C., and Beil, D. G. (**1979**). ''Hair cell condition and auditory nerve response in normal and noise-damaged cochleas,'' Acta Oto-Laryngol. **88**, 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.

Martin, G. K., Foster, P. K., Stagner, B. B., and Lonsbury-Martin, B. L. (**1998a**). ''Effects of octave-band noise exposure on DPOAE suppression-tuning curves in rabbits,'' Abstr. Assoc. Res. Otolaryngol. **21**, 598.

Martin, G. K., Jassir, D., Stagner, B. B., and Lonsbury-Martin, B. L. (**1998b**). ''Effects of loop diurectics on the suppression tuning of distortion-product otoacoustic emissions in rabbits,'' J. Acoust. Soc. Am. **104**, 972–983.

Martin, G. K., Stagner, B. B., Jassir, D., Telischi, F. F., and Lonsbury-Martin, B. L. (**1999**). ''Suppression and enhancement of distortion-product otoacoustic emissions by ITs above f_2. I. Basic findings in rabbits,'' Hear. Res. **136**, 105–123.

Mills, D. M. (**1998**). ''Interpretation of distortion product otoacoustic emission measurements. II. Estimating tuning characteristics using three stimulus tones,'' J. Acoust. Soc. Am. **103**, 507–523.

Mills, D. M. (**2000**). ''Frequency responses of two- and three-tone distortion product otoacoustic emissions in Mongolian gerbils,'' J. Acoust. Soc. Am. **107**, 2586–2602.

Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (**1998**). ''Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae,'' Science **282**, 1882–1884.

Pienkowski, M., and Kunov, H. (**2001**). ''Suppression of distortion product otoacoustic emissions and hearing threshold,'' J. Acoust. Soc. Am. **109**, 1496–1502.

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.

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.

Rubsamen, R., Mills, D. M., and Rubel, E. W. (**1995**). ''Effects of furosemide on distortion product otoacoustic emissions and on neuronal responses in the anteroventral cochlear nucleus,'' J. Neurophysiol. **74**, 1628–1638.

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.

Salvi, R. J., Hamernik, R. P., and Henderson, D. (**1978**). ''Discharge patterns in the cochlear nucleus of the chinchilla following noise induced asymptotic threshold shift,'' Exp. Brain Res. **32**, 301–320.

Salvi, R. J., Henderson, D., Hamernik, R., and Parkins, C. (**1980**). ''VIII nerve response to click stimuli in normal and pathological cochleas,'' Hear. Res. **2**, 335–342.

Sewell, W. F. (**1984**). ''The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats,'' Hear. Res. **14**, 305–314.

Slepecky, N. (**1986**). ''Overview of mechanical damage to the inner ear: Noise as a tool to probe cochlear function,'' Hear. Res. **22**, 307–321.

Walsh, E. J., and McGee, J. (**1990**). ''Frequency selectivity in the auditory periphery: Similarities between damaged and developing ears,'' Am. J. Otolaryngol. **11**, 23–32.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L_2 below L_1,'' J. Acoust. Soc. Am. **97**, 2346–2358.

Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects

Van Summers
Army Audiology & Speech Center, Walter Reed Army Medical Center, Washington, DC 20307-5001

Egbert de Boer
Room D2-226, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Alfred L. Nuttall
Oregon Hearing Research Center, NRC04, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201-3098 and Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506

(Received 27 November 2002; accepted for publication 21 April 2003)

Harmonic complexes comprised of the same spectral components in either positive-Schroeder (+Schr) or negative-Schroeder (−Schr) phase [see Schroeder, IEEE Trans. Inf. Theory **16**, 85–89 (1970)] have identical long-term spectra and similar waveform envelopes. However, localized patterns of basilar-membrane (BM) excitation can be quite different in response to these two stimuli. Measurements in chinchillas showed more modulated (peakier) BM excitation for +Schr than −Schr complexes [Recio and Rhode, J. Acoust. Soc. Am. **108**, 2281–2298 (2000)]. In the current study, laser velocimetry was used to examine BM responses at a location tuned to approximately 17 kHz in the basal turn of the guinea-pig cochlea, for +Schr and −Schr complexes with a 203-Hz fundamental frequency and including 101 equal-amplitude components from 2031 to 22 344 Hz. At 35-dB SPL, +Schr response waveforms showed greater amplitude modulation than −Schr responses. With increasing stimulation level, internal modulation decreased for both complexes. To understand the observed phenomena quantitatively, responses were *predicted* on the basis of a linearized model of the cochlea. Prediction was based on an ''indirect impulse response'' measured in the *same* animal. Response waveforms for Schroeder-phase signals were accurately predicted, provided that the level of the indirect impulse used in prediction closely matched the level of the Schroeder-phase stimulus. This result confirms that the underlying model, which originally was developed for noise stimuli, is valid for stimuli that produce completely different response waveforms. Moreover, it justifies explanation of cochlear filtering (i.e., differential treatment of different frequencies) in terms of a linear system.
[DOI: 10.1121/1.1580813]

PACS numbers: 43.64.Kc, 43.64.Bt [WPS]

I. INTRODUCTION

Perception and discrimination of auditory stimuli can be studied with a great variety of signals. Commonly used classes of stimuli include (this compilation is not at all exhaustive) tone signals, noise signals, impulsive signals, and multicomponent signals. A similar variety of stimuli can be applied in research into cochlear function. To facilitate understanding auditory processes, the stimulus signals used in both psychophysical and physiological studies are often ''synthetic'', i.e., not recorded from natural sounds but generated for a specific purpose.

In most cases of observing responses of the auditory system to these stimuli, *nonlinear* phenomena in the auditory system play a part, and one goal of hearing research is to bring unity to the description of these effects. In this respect the methodologies of psychophysics and physiological research have generally been very different. One of the purposes of the present study is to attempt to bring these fields closer together. To this aim, very specific stimuli have been used in experiments on movements of the basilar membrane (BM) in test animals, and the results can be used to try to understand data of psychophysical experiments in humans and birds in which the same stimuli were used—but in a different setting, namely, one involving auditory masking.

Two important properties of cochlear mechanical nonlinearity are that response magnitudes show compressive growth and reduced frequency selectivity as input levels increase. This applies to all classes of stimuli listed above. Surprisingly, these nonlinear changes in magnitude can, in specific cases, have remarkably little effect on the timing properties of the response. Responses of the BM to impulse stimuli, measured at a fixed basal location *tuned to a high frequency* (e.g., 17 kHz), show almost the same zero-crossing times for various levels of stimulation. Concurrently, the trajectories of the instantaneous frequency with time (''glides'') follow the same pattern over a large (>80-dB) input range (de Boer and Nuttall, 1997, 2000; Recio *et al.*, 1998; Recio and Rhode, 2000). The upward frequency glides in the high-frequency region reflect dispersion in the traveling wave, whereby lower-frequency components of the

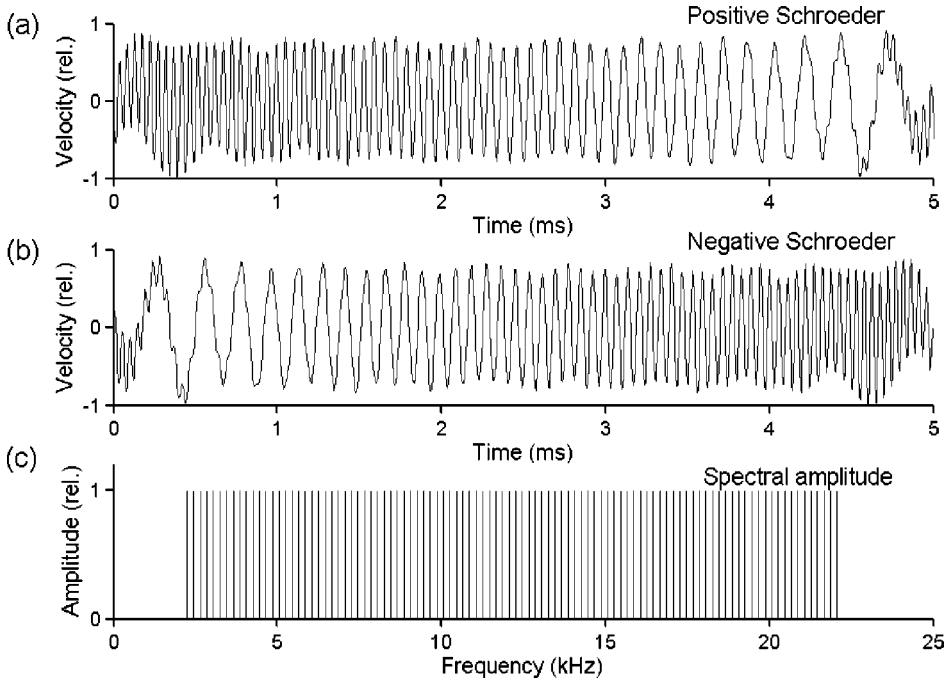


FIG. 1. Schroeder-phase signals. Panel (a): Positive Schroeder waveform, shifted over 4.5 ms with respect to origin of signal buffer. Panel (b): Negative Schroeder wave, shifted over 0.5 ms with respect to origin of signal buffer. Panel (c): Amplitude spectrum.

stimulus are traveling faster and arrive at the measurement location sooner than higher frequencies (Pickles, 1988; Shera, 2001). In different terms, the glide phenomenon is related to negative curvature of response phase with increasing frequency. Similar frequency glides are present in derived impulses responses of auditory-nerve fibers in basal regions (Møller and Nilsson, 1979; Carney *et al.*, 1999).

In many listening situations these intrinsic frequency glides in BM response are of little perceptual significance. However, psychophysical results involving ''Schroeder-phase'' harmonic complexes indicate an important influence of this dispersion on auditory processing (Smith *et al.*, 1986; Kohlrausch and Sander, 1995). These complex stimulus signals are periodic and contain a finite number of equal-amplitude harmonics with phases θ_n (with $n=1,2,3,...$) based on the following formula, proposed by Schroeder (1970):

$$\theta_n = \pm \pi n(n-1)/N_c, \tag{1}$$

where θ_n is the phase in radians of the nth harmonic and N_c is the total number of harmonics. The quadratic term in Eq. (1) causes a non-negative second derivative of the phase; in the sequel we will refer to *phase curvature* as directly related to this second derivative. Waveforms of the two Schroeder-phase signals used in the present study are plotted in Fig. 1, panels (a) and (b); see Sec. II for more details. Note that the waveform amplitudes are almost constant. The legend describes how much the two waveforms are shifted with respect to the ''zero point'' in time. The positive-Schroeder (in what follows, designated by ''+Schr'') complex is produced with Eq. (1) with the ''+'' sign and each period is observed to contain a high-to-low frequency glide. The negative-Schroeder (designated by ''−Schr'') complex is produced with the ''−'' sign in Eq. (1), and it is a time-reversed version of the +Schr complex, each period containing a low-to-high frequency glide. Panel (c) of Fig. 1 shows the amplitude spectrum of the Schroeder signals. The Appendix describes a fundamental problem in interpreting and unwrapping phase values in terms of Eq. (1) when signals—stimuli as well as responses—are given or acquired as *waveforms*. The Appendix illustrates that the second derivative is preserved in the unwrapped phase, a most important property in the context of this paper. It also contains an introduction into the nature of responses to Schroeder-type signals.

Psychophysically, +Schr stimuli are often found to be less effective maskers than −Schr complexes composed of the same harmonic components (Smith *et al.*, 1986; Kohlrausch and Sander, 1995; Carlyon and Datta, 1997a,b; Summers and Leek, 1998; Summers, 2000; Oxenham and Dau, 2001a,b). In effect, pure-tone masked thresholds may differ by more than 20-dB between +Schr and −Schr maskers with identical (long-term) magnitude spectra. The common interpretation of these masking differences is that the upward frequency glide that is inherent in the BM impulse response partially compensates for the downward glide in the +Schr complex to produce a more modulated (''peakier'') internal response pattern for +Schr than −Schr stimuli. These high-amplitude peaks in the response to the +Schr complex alternate with low-amplitude (''quiet'') periods, where the presence of an added target signal is readily detected. This results in reduced masking effectiveness for the +Schr complex. A recent physiological study has provided support for this difference in modulation of the BM response to Schroeder signals: Recio and Rhode (2000) reported more modulated (peakier) BM excitation for +Schr than −Schr complexes in chinchillas.

In the psychophysical studies the difference in masking effectiveness between +Schr versus −Schr complexes varies with level (Carlyon and Datta, 1997a,b; Summers and Leek, 1998; Summers, 2000). This may indicate that the relative peakiness of BM responses produced by +Schr and −Schr complexes may also be level dependent. A number of level-dependent changes in BM response properties may influence

the peakiness of responses to these stimuli. In effect, we know much about how stimulation level influences BM response properties. With increasing level, the low-frequency tail of the frequency response at a fixed BM location grows linearly, while responses near the *characteristic frequency* (CF) grow less than linearly (Rhode, 1971; Johnstone *et al.*, 1986; Ruggero *et al.*, 1997). As a result, a sharply-tuned bandpass response at low levels gradually turns into a broadly tuned, more or less low-pass response as level increases. In addition, the frequency producing the largest response at a fixed basal BM location (the *best frequency*, BF) shifts downward with increasing level. Both of these changes lead to an increased influence of frequencies well below CF on BM responses at high levels (relative to frequencies in the CF region). These level-dependent changes in the magnitude response could enhance or reduce peakiness in BM response envelopes depending on the relative influence of low- versus high-frequency components in controlling response modulation. In addition, although the temporal properties of BM impulse responses are nearly invariant with level, phase responses to tones and clicks do show some level dependence. For frequencies just below CF, responses to low-level stimuli show phase leads relative to higher-intensity inputs. Just above CF, the pattern is reversed and low-level inputs show phase lags relative to higher-level signals (Rhode and Robles, 1974; Geisler and Rhode, 1982; Sellick *et al.*, 1982; Nuttall and Dolan, 1996; Ruggero *et al.*, 1997; Recio *et al.*, 1998). These differences indicate level-dependent changes in the temporal responses to near-CF frequencies which could also have significant effects on BM response envelopes for a range of signals.

In the present study we examined how presentation level influences BM responses to +Schr and −Schr stimuli at a fixed BM location in guinea pigs. The observed responses were put into relation with *impulse responses* of the BM at comparable stimulus levels. The impulse response of the BM can be found in two ways.

(i) With an acoustic impulse as stimulus (called the ''direct'' impulse response); and
(ii) With a wideband flat-spectrum noise stimulus (called the ''indirect'' impulse response). It has been shown by de Boer (1997) that, in view of an extremely general type of *nonlinear model of the cochlea*, the ''indirect'' impulse response as determined by method (ii) is to be preferred for analysis. The indirect impulse response is found from the cross-correlation function of input and output signals, and it is this cross-correlation function that shows the characteristic ''peakiness'' that we associate with an ''impulse response'' (de Boer and Nuttall, 1997).

In a linear system, direct and indirect impulse responses are identical. In that case, the impulse response will allow accurate prediction of responses to other inputs such as tones and harmonic complexes. Despite nonlinearity in BM response properties, measured direct impulse responses have shown fairly good agreement with indirect impulse responses (de Boer and Nuttall, 1997). The same was found for direct impulse responses computed from BM responses to tones (Robles *et al.*, 1976; Recio and Rhode, 2000). In addition, responses to pure tones can be predicted very well from (indirect) impulse responses (de Boer and Nuttall, 2002). It is the average amplitude of the BM response that is the prime determinant of nonlinearity (de Boer and Nuttall, 2000); hence, levels have to be matched for optimum results. As said earlier, we are studying the relation between responses to Schroeder stimuli and (indirect) impulse responses. In point of fact, we will try to predict responses from stimulus signals and impulse responses. If that prediction succeeds, we have the means to explain all our observations and to extrapolate these and other findings to psychophysics.

II. DATA ACQUISITION AND PROCESSING

A laser velocimeter was used to examine BM movements in the basal turn of the cochlea for eight pigmented guinea pigs. The preparation, measurement, and data-processing techniques were as described previously by de Boer and Nuttall (1997). Velocities were measured at a location tuned to approximately 17–17.5 kHz at low levels. The custom-designed data presentation and collection hard- and software operated at a clock frequency of 208 kHz and continuously presented a (circular) 4096-point data array while simultaneously sampling the BM response. The period of the generated stimulus as well as of the response signals was ~20 ms. Test stimuli were +Schr and −Schr complexes with a fundamental frequency of ~203.1 Hz and were the sum of 101 equal-amplitude components between 2031 and 22 344 Hz; see Fig. 1, panel (c). These 101 components were assigned phase values based on the ''+'' or ''−'' versions of Eq. (1) to produce the waveforms plotted in Fig. 1, panels (a) and (b). The complete 4096-point data array contained four periods of the waveform, each corresponding to ~5 ms. Velocity responses to pseudorandom noise stimuli and to +Schr and −Schr complexes were collected over a range of stimulus levels. Schroeder stimuli were presented at 35, 55, 75, and 95 dB SPL, and noise stimuli at levels from 10 to 100 dB SPL, in steps of 10 dB. All these levels refer to the power contained in one octave from 10 to 20 kHz.[1] Stimuli were calibrated with a $\frac{1}{4}$-in. condenser microphone placed in an ultrasmall coupler. In earlier experiments we confirmed the calibration *in situ* with a probe microphone at the entrance of the bony auditory meatus. Analysis and processing of data were done with MATLAB® in conjunction with routines written in TURBO PASCAL®.

III. RESULTS

A. "Indirect" impulse responses

Panel (a) of Fig. 2 shows waveforms of ''indirect'' impulse responses of the BM for one animal at levels from 40 to 100 dB SPL. The curves have been normalized to the same rms amplitude to facilitate comparison of the temporal waveforms. Responses have amplitudes that first grow and later decrease with increasing time. As stimulus level increases, the later portions of the response become smaller in comparison to earlier regions. The zero crossings indicate an oscillation frequency that increases with time (the ''glide''). Note the temporal alignment of the oscillations across levels:

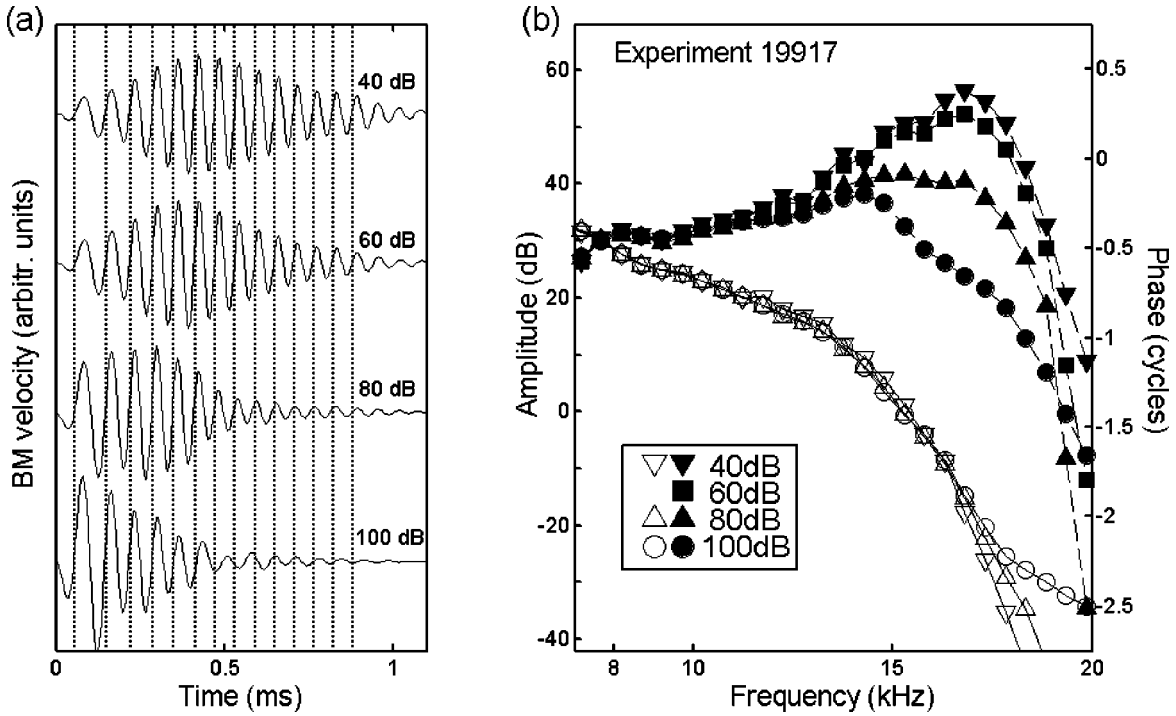


FIG. 2. Characteristics of ''indirect impulse responses.'' Panel (a): Waveforms for various stimulus levels, normalized *re*: rms amplitude. Measured responses are corrected for the stapes response. Input levels are in dB SPL for the 10–20 kHz octave. Curves are displaced vertically with respect to one another. Panel (b): Amplitude and phase spectra corresponding to the waveforms. Filled symbols show response amplitudes and depict ''gain'' since data have been normalized to the stapes velocity at the same stimulus level. Unfilled symbols indicate phase. The 60-dB phase curve was nearly identical to 40-dB data and is not included in panel (b). Note that a linear frequency scale is used in panel (b) to accurately represent the curvature in phase in the CF region.

the locations of the zero crossings are nearly invariant with level (see the vertical dotted lines). These temporal patterns are representative of all eight animals.

Panel (b) of the figure shows amplitude and phase values based on Fourier transforms of the waveforms in panel (a) Spectral amplitude values have been normalized with respect to the stapes response, *taken at the same stimulus level* so that the ordinate depicts cochlear gain. The amplitude of the frequency response shows fairly narrow tuning at low levels with a characteristic frequency (CF) of about 17 kHz. As reported in many previous studies, the amplitude curves show ''tip'' and ''tail'' regions which respond differently to level variation. The boundary between these regions is about half an octave below CF, i.e., around 12 to 13 kHz. For low frequencies, the response is linear with stimulus amplitude, and, consequently, the curves overlap. At higher levels the response for high frequencies is compressive (it varies less than linearly with stimulus amplitude). As the level increases, the amplitude in the tip region thus diminishes relative to that at lower frequencies, tuning broadens, and the amplitude peak (at the *best frequency*, or BF) moves to a lower frequency.

In contrast to the amplitude curves, the phase curves are fairly similar across stimulus levels. In detail, the phase curves also show evidence of different patterns in tip and tail regions. In the tail region (below around 13 kHz), phase change is approximately linear with frequency, signifying a constant delay. Here, the phase curves for different stimulus levels almost perfectly overlap. At higher frequencies, phase values have an increasing negative slope (i.e., negative curvature) with increasing frequency. This indicates that greater delays are associated with high frequencies than low ones. In the CF region the phase curves show small, but systematic, level-dependent changes as detailed in the Introduction. Note that a linear frequency scale is used in panel (b) to accurately represent the curvature in phase in the CF region. Wave dispersion, which is the cause of frequency-to-place mapping in the cochlea, shows up in the form of phase curvature—on both a logarithmic and a linear frequency scale.

B. Responses to harmonic complexes

Figure 3 shows typical response waveforms, measured at a medium stimulus level (55 dB), for a +Schr and a −Schr stimulus for one animal. The abscissa covers the most interesting part of the response, over 3 ms. The legend tells how much the waveforms are shifted in time. The data are normalized to the same rms value. The 3-ms temporal interval allows inspection of the frequency glides with time. The +Schr response (upper waveform) has a high-to-low glide, in the same direction as the stimulus signal [compare Fig. 1(a)]. In contrast, the response to the −Schr stimulus (lower waveform) has a low-to-high glide [compare Fig. 1(b)].

Figure 4 shows data in a form emphasizing the *envelope*, over a range of stimulus intensities. The envelope was determined from an expansion to the *analytic signal* of all waveforms. In each panel the responses are shown normalized to the same rms value The left-hand and middle columns show BM responses to the two harmonic complexes at stimulus levels of 35, 55, 75, and 95 dB SPL. The right-hand column shows indirect impulse responses at similar levels. In each panel of the figure, one full period (~5 ms) is shown.

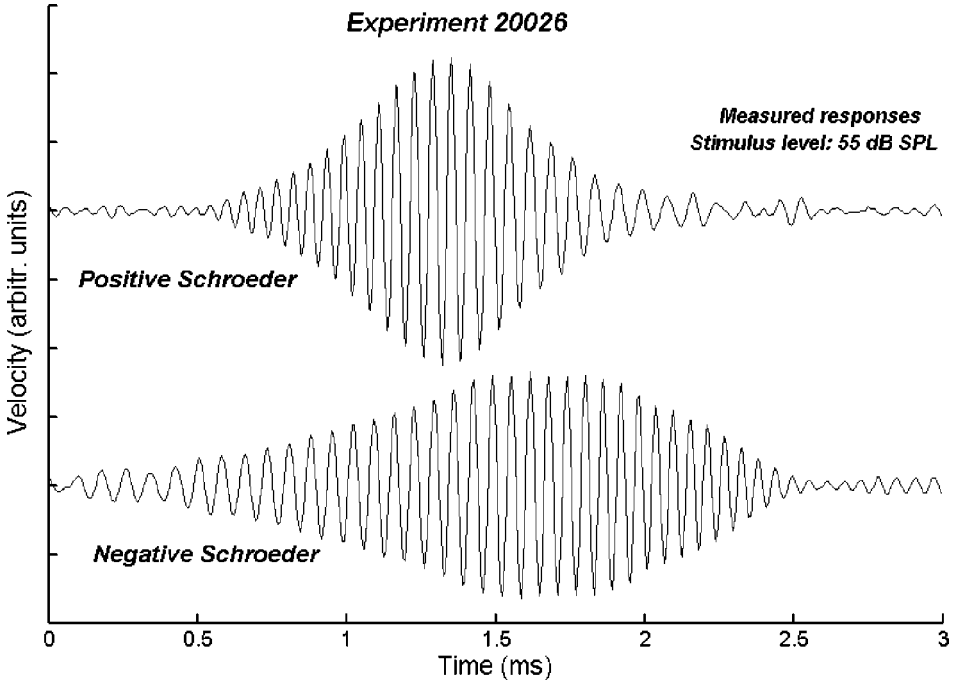


FIG. 3. Examples of response waveforms for Schroeder-phase stimuli. Stimulus level: 55 dB SPL in the octave from 10 to 20 kHz. Responses normalized to equivalent rms amplitude. Positive Schroeder response is plotted above negative Schroeder response, which has been shifted over 3 ms to the left.

Indirect impulses responses become peakier and response peaks move earlier with increasing level. These changes are consistent with a greater influence of low-frequency parts of the input on BM responses at high levels. (see the Introduction and Fig. 2). That is, given that low frequencies propagate more rapidly and less dispersively than higher frequencies, peakier and earlier impulse responses should occur with increasing level as low-frequency portions of the impulse come to dominate the internal response.

At low levels, +Schr stimuli produce peakier BM response patterns than −Schr stimuli. As level increases, both responses are showing less and less peakiness, but the −Schr response shows this effect more clearly. These results are consistent with previously reported BM displacement data for chinchillas (Recio and Rhode, 2000). Although response envelopes for +Schr and −Schr stimuli differ (at low input levels in particular), responses are similar in overall energy. That is, rms values of responses to the two stimuli (at a constant input level) are nearly equivalent. This differs from the earlier work involving chinchillas where responses to −Schr stimuli had higher rms values (Recio and Rhode, 2000).

The stimulus- and level-dependent changes in BM response envelopes shown in Figs. 3 and 4 were present for all eight animals. Figure 5 shows results for another animal, organized in the same way as Fig. 4. As can be seen in both figures, responses to the Schroeder-phase stimuli often have

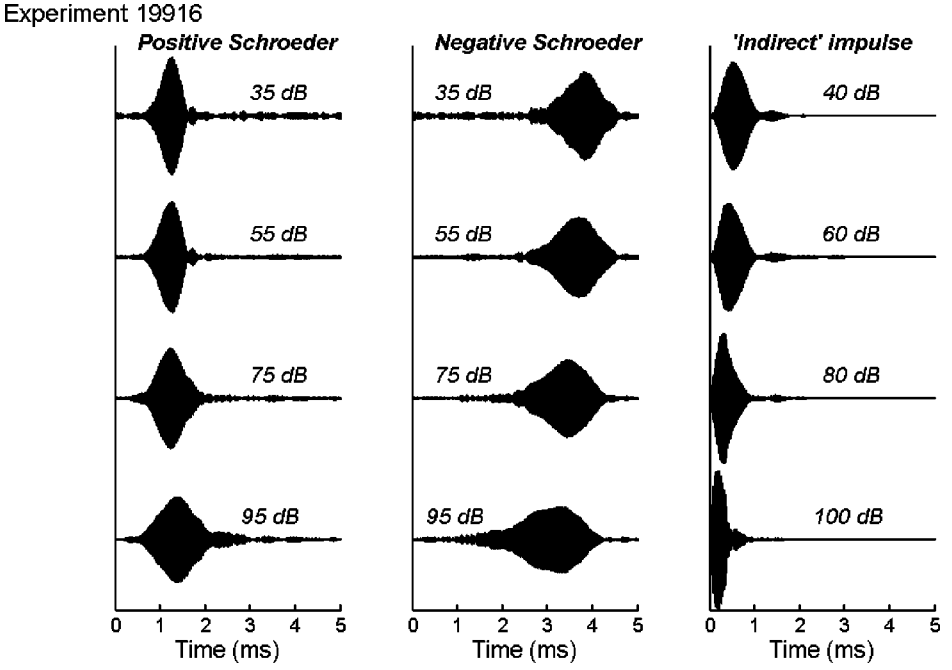


FIG. 4. Envelopes of responses to Schroeder signals and indirect impulse responses. Left panel: responses to positive Schroeder (+Schr) signals, for four different stimulus levels. Center panel: responses to negative Schroeder (−Schr) signals, for the same levels. These responses have been shifted by 4.25 ms to the left. Right panel: indirect impulse responses at four levels. Note that the responses become *wider* with increasing stimulus level for the two Schroeder-phase signals while the impulse response narrows. In addition, the +Schr response shows a steady increase in the latency of its peak as level increases while, for both the −Schr and the indirect impulse response, response peaks move earlier with level (see the text).

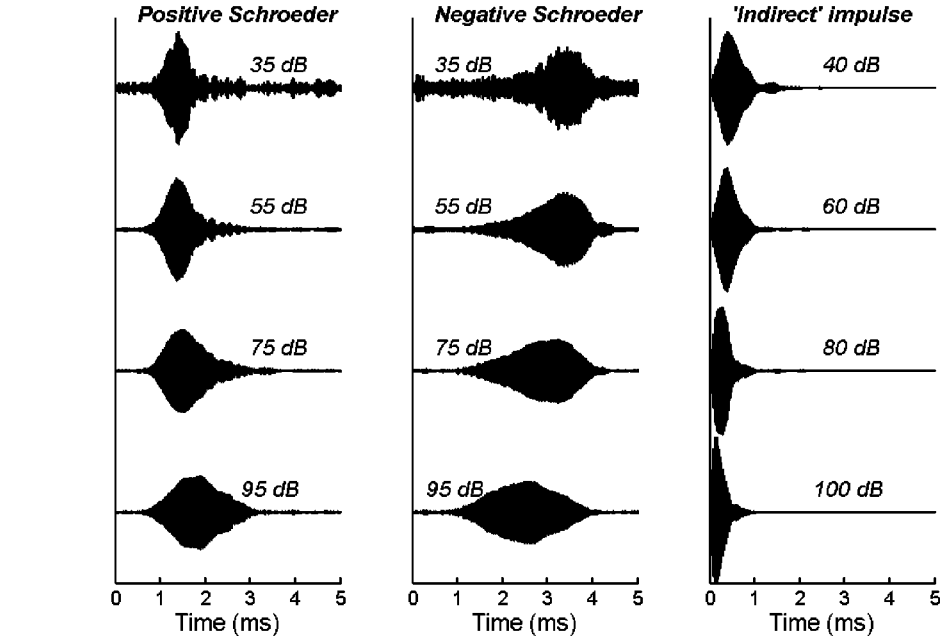


FIG. 5. As Fig. 4, for a different experiment.

the form of ''spindles'' of which the slowly decaying part points to the right (i.e., follows the response peak) for responses to +Schr signals, and to the left (preceding the peak) for responses to −Schr signals. Note that the relative amplitude of this slowly decaying part of the response steadily increases with input level, particularly for the −Schr stimulus. Thus, with increasing level, the response to the −Schr stimulus is more and more ''filled up'' within a period. Responses to the +Schr stimuli also show some of this ''filling'' with level, but more nearly retain their low-level pattern, a peak region followed by a long period of ''silence.'' Input level also influences the temporal locations of the response peaks. Similar to the pattern just described, the peak moves to the right for +Schr stimuli and to the left for −Schr complexes.

The level-dependent changes in the response patterns for the +Schr and −Schr stimuli (peak shifts and filling on one side of the peak) are again consistent with a greater influence of the low-frequency parts of the input on responses as levels increase (recall that the low-frequency parts of the stimulus signal occur late within each period for the +Schr stimuli and early for the −Schr complexes). The peak shifts and partial filling of the response period with level are among the effects that should be captured by the level-dependent changes in the indirect impulse responses (Fig. 2) in order for predictions based on the impulse data to provide a close match to measured responses for the Schroeder-phase stimuli. This modeling is described next.

IV. PREDICTION OF RESPONSES TO HARMONIC STIMULI

In a linear system, the impulse response is invariant with input level and allows accurate prediction of responses to arbitrary stimuli. In the nonlinear cochlea, indirect impulse responses show large changes in magnitude spectra and small changes in phase with level, and may not necessarily predict responses to other inputs well. To examine the accuracy of prediction, *measured* responses to +Schr and −Schr stimuli are to be compared to *predicted* responses based on the indirect impulse response data. Predicted responses are determined by filtering the stimulus waveforms using filter characteristics derived from indirect impulse responses. The first issue to address is to assess *the most appropriate stimulus level* for the impulse response. Let us assume that in the cochlea, outer hair cells (OHCs) produce a place-frequency-specific enhancement of the response and that physiologic saturation in OHCs is the (sole) cause of cochlear nonlinearity. In de Boer and Nuttall (2000), it is shown that the shape of the indirect impulse response is fairly accurately determined by the *average degree of saturation* of OHCs for the noise stimulus employed. In the paper cited, the correspondence is shown for noise stimuli and indirect impulse responses, in fact, for stimuli and responses that are regular in the stochastic sense, and certainly not peaky in their waveforms. For such stimuli the average degree of saturation directly depends on the level of the stimulus, and on nothing else. The close correspondence between direct and indirect impulse responses, mentioned in the Introduction, indicates that the stimulus or the response being peaky or nonpeaky is not an important condition. In a recent paper, de Boer and Nuttall (2002) have shown that the correspondence is also true for tones versus noise. That the stimulus has a narrow spectrum is, therefore, not a constraint either. For all these reasons, we may expect that response prediction will also work in the case of Schroeder stimuli.

Prediction is carried out by convolving the stimulus waveform with the appropriate indirect impulse response. A point to be noted is that our indirect impulse responses are compensated for the response of the stapes. The same has to be done for the measured responses to Schroeder stimuli, otherwise, the prediction will not work. The upper portion of Fig. 6 shows *measured* and *predicted* response waveforms for a 35 dB SPL +Schr stimulus and the lower part of the

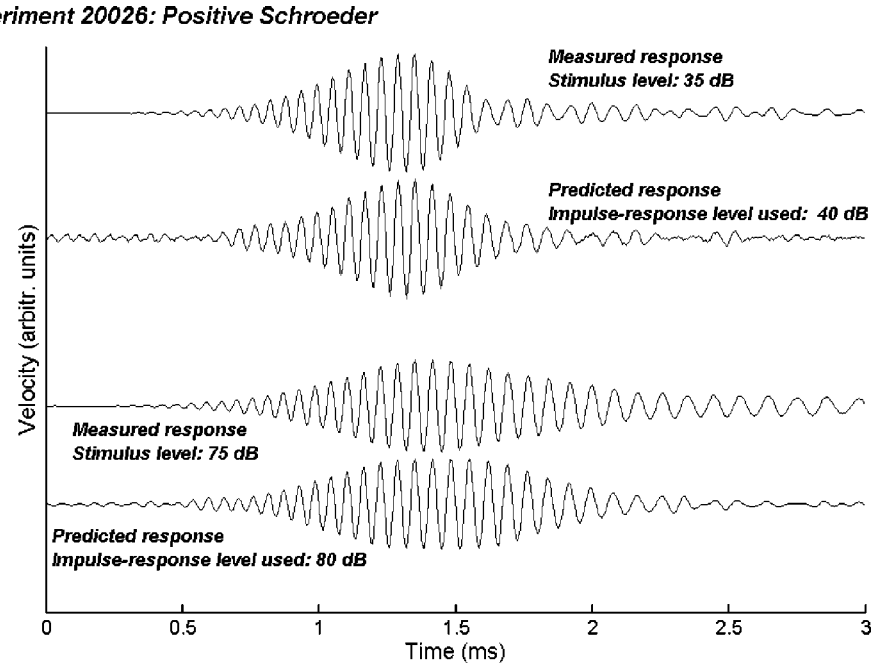


FIG. 6. Typical measured BM responses to a positive Schroeder stimulus and predicted response waveform for that signal based on indirect impulse responses. Upper waveform pair shows measured and predicted responses at low stimulus level; lower pair shows high stimulus level. Note the fairly accurate reproduction of envelope waveform and frequency glide.

figure shows the same data for a 75 dB SPL stimulus. The presentation levels for Schroeder stimuli and noise are closely matched, assuming that this leads to a similar average degree of saturation. Note that the abscissa scale has been shortened to go from 0 to 3 ms. It is clear that measured and predicted responses are very similar: Pearson correlation coefficients between measured and predicted waveforms for both the 35- and 75-dB responses are approximately 0.96.

Figure 6 shows details of both envelope and fine structure of the measured and predicted waveforms. A Lissajous figure, in which all values from the predicted response waveform are plotted against corresponding values in the measured response, gives a second view of the relation between measured and predicted signals. Figure 7 shows a set of such Lissajous figures. In the vertical direction, the stimulus level in the panels varies from 35 to 95 dB SPL, in steps of 20 dB. In the horizontal direction it is the stimulus level of the indirect impulse response with which the response is predicted that is varying, and this level varies in steps of 10 dB. Along two of the rising diagonals of the figure (the one beginning in the lower left-hand corner and the one immediately to its right) the impulse level used in prediction differs by only 5 dB from the actual response level. The remaining diagonals involve differences of 15, 25, or 35 dB between the indirect impulse (predictor) and the response level. In each panel, waveform values from actual responses are plotted as a function of predicted values; a 45-degree line within a given panel would represent perfect prediction. Pearson r values correlating predicted and actual values are shown in each panel. In general, correlations were high (>0.9) when pre-

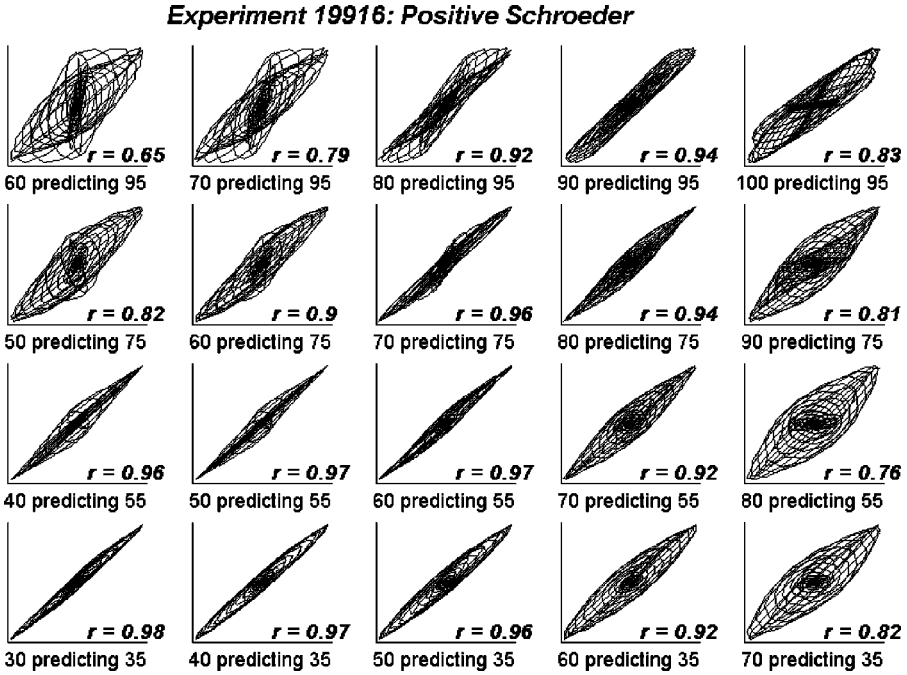


FIG. 7. (Experiment 19916). Measured and predicted responses for *positive* Schroeder stimuli. Lissajous figures of predicted response (horizontal) versus measured response (vertical). All signals are normalized with respect to 1.05 times their maximal values and the scales go from -1 to $+1$. In one row, predictions are done with different stimulus levels of the (indirect) impulse response, and the stimulus level of the response is constant. The abscissa label of every panel tells the level of the impulse response used for prediction, and the level of the response to be predicted. All levels are in dB SPL between 10 and 20 kHz.

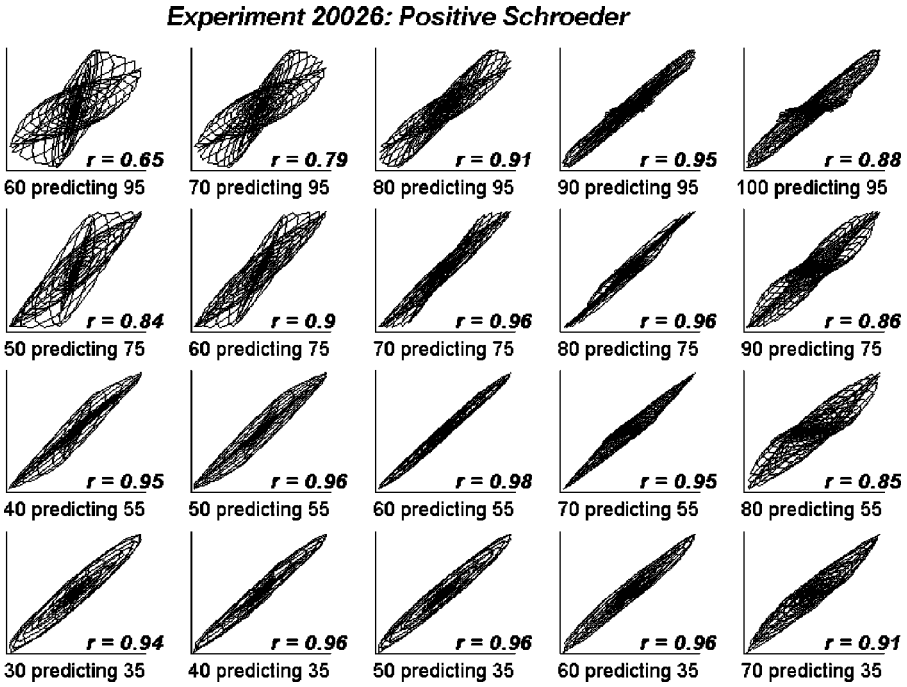


FIG. 8. (Experiment 20026). Measured and predicted responses to positive Schroeder stimuli. Layout as Fig. 7.

dictions and measured responses were closely matched in level, and were found to decrease steadily as the difference in level between the impulse and harmonic stimuli increased. As suggested by the high Pearson correlation coefficients along the two central diagonals of Fig. 7, and as directly indicated by the measured and predicted responses shown in Fig. 6, the prediction procedure accurately recovered both the peak shifts and increased filling of the response period with increasing level.

It should be realized that a correlation coefficient is not necessarily the best descriptor of the similarity in shape of two waveforms because it intrinsically weighs the larger amplitudes quadratically. Obviously, the variations in response to Schroeder-type stimuli are complex and subtle. On the other hand, a ''narrow'' Lissajous figure means that variations in envelope are simulated well. We also verified that, wherever the Lissajous figures are narrow, the intrinsic frequency modulation of the waveform (the glide) is well predicted. This means that the fine structure is preserved, too. Figure 8 shows Lissajous figures for another experiment. Across experiments, Lissajous figures tended to show the best correspondence when predictor and response levels differed by only 5 dB. Figures 7 and 8 show predictions and measured responses to +Schr stimuli. Figure 9 shows an example of predictions for −Schr signals. Since it is similar to the earlier figures and does not lead to different conclu-

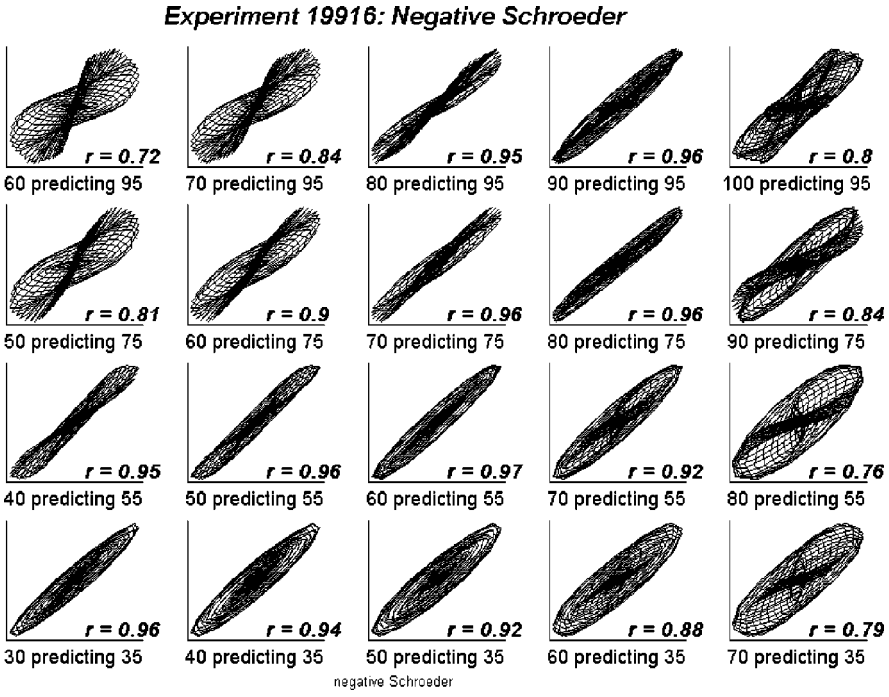


FIG. 9. (Experiment 19916). Measured and predicted responses for negative Schroeder stimuli. Layout as Fig. 7.

sions, we refrain from presenting figures for −Schr signals for other experiments.

V. DISCUSSION

In the present study, BM velocity response envelopes in guinea pigs were peakier for +Schr than −Schr complexes, particularly at low input levels (see Figs. 4 and 5). As described earlier, peaky responses to +Schr complexes are consistent with (partial) cancellation of the high-to-low frequency glide of the +Schr signal (or equivalently, positive curvature in the stimulus phase structure) by the low-to-high frequency glide (or negative curvature in response phase) inherent to basal cochlear mechanics. The indirect impulse responses allowed accurate prediction of responses to the two Schroeder-phase complexes over a 60 dB input range. The results of this analysis are discussed more fully below, along with a consideration of how spatial dispersion of frequencies during traveling-wave propagation may account for certain aspects of the data, and some implications of these data for previously reported psychophysical results.

A. Predicting response to Schroeder-phase complexes based on indirect impulse data

At low stimulation levels, prediction of responses is straightforward since we may assume the cochlea to be operating linearly. We have employed exactly the same technique for prediction at higher stimulus levels, with the only modification that the cochlear transfer function is taken from the indirect impulse response determined at an appropriate stimulus level. The predictions proved to be quite good, at least when levels were closely matched. Lissajous figures demonstrate the agreement clearly. We verified in a separate test that not only the envelope modulation of the response is predicted well but also the fine structure. The reason why Lissajous figures are not all like straight lines probably lies in slow variations of the phase that are not captured in the prediction. The nonlinear processes in the cochlea apparently do not contain subprocesses that are strongly frequency dependent, or are disrupting temporal relations, and we are fully justified in continuing an analysis as if the cochlea is linear at any one level as we employed in the experiments. The fundamental concept underlying the use of noise stimuli and spectral analysis in a nonlinear system as the cochlea lies in the EQ-NL theorem (de Boer, 1997). This theorem prescribes the conditions to be met by the system for this quasilinear analysis to be justified. As far as we know, the cochlea can well be modeled by such a system.

Consider Fig. 11 in the Appendix. Thick lines in the two panels of the figure show the course of the (unwrapped) phase for a positive and a negative Schroeder stimulus, respectively. The figure shows that the sampling of frequencies and the unwrapping of phase that have taken place do not disrupt the size and polarity of the phase curvature (the second derivative). The phase of the cochlear transfer function has to be added to the phase of the stimulus and this summed phase function always has a negative curvature. That curvature is largest in the frequency region of the response peak. Hence, its effect on the response will be large whenever that frequency region is important. In other words, at low stimulus levels, where the response has a sharp spectrum centered at the CF, the phase modification will be large. This situation is depicted by the thin curve in the upper panel of Fig. 11, the positive phase curvature is *diminished* locally, i.e., in the frequency region of the peak (the thick bar on the abscissa).

For −Schr stimuli (lower panel, Fig. 11) at low stimulus levels the effect on the response phase is to *enhance* the curvature in that frequency region. Figure 12 of the Appendix shows prototype waveforms for these two cases. In both figures the curvature of the phase is taken as constant. In the upper waveform it is small and positive, and in the lower waveform, large and negative. In the construction of that figure the amplitude spectra were stylized (see the legend); they were not meant to correspond to cochlear responses. Nevertheless, an important property can be deduced from the figure, a property that can also be discerned in actual responses. That property is the following: signals with a restricted spectrum and a *large* phase curvature have a *long* waveform (a long ''spindle''). Conversely, signals with a *small* phase curvature have a *short* waveform.

At higher stimulus levels the region of large phase curvature in the response becomes less important so that the curvature due to the Schroeder stimulus signal dominates the response phase. Consequently, the response waveforms tend to show the character of the prototype response of the lower waveform in Fig. 12. The fact that the actual spectrum of the response is not the stylized one used for that particular figure causes the spindles of both the measured and predicted responses to deviate from the prototypes. Hence, the pronounced effects of stimulus level. In particular, with increasing level, BM responses decrease less in modulation for −Schr than for +Schr stimuli. This reasoning also shows why the predictions depicted by the Lissajous figures fail when the level difference is large. In that case the weighting of the spectral components is way off, and a good prediction cannot be reached. The Lissajous procedure is very sensitive in this respect.

Actually, even at low levels the compensation of the Schroeder phase curvature is not complete because the narrowest indirect impulse responses in Figs. 4 and 5 are peakier than the +Schr responses. It may be asked: are these indirect impulse responses the peakiest possible responses in the cochlea? The answer is no. It is possible to construct a hypothetical stimulus that precisely compensates for the cochlear glide. For this purpose the *amplitude* frequency spectrum of a Schroeder stimulus is taken as the spectral amplitude [see Fig. 1(c)]. This amplitude is endowed with the opposite of the phase spectrum of the indirect impulse response. The resulting signal is noncausal (i.e., nonzero portions of the signal precede time zero). If such a signal could be presented to the cochlea, the response (also noncausal) would have a constant phase for all component frequencies involved. Figure 10 shows a typical high-level indirect impulse response and the predicted, *artificial response* to the hypothetical stimulus. The latter response is the peakiest that could ever be obtained in this cochlear region, given the amplitudes in the frequency spectrum.[2] The effective duration of even the peakiest impulse response of the cochlea is larger than the duration of this hypothetical response.

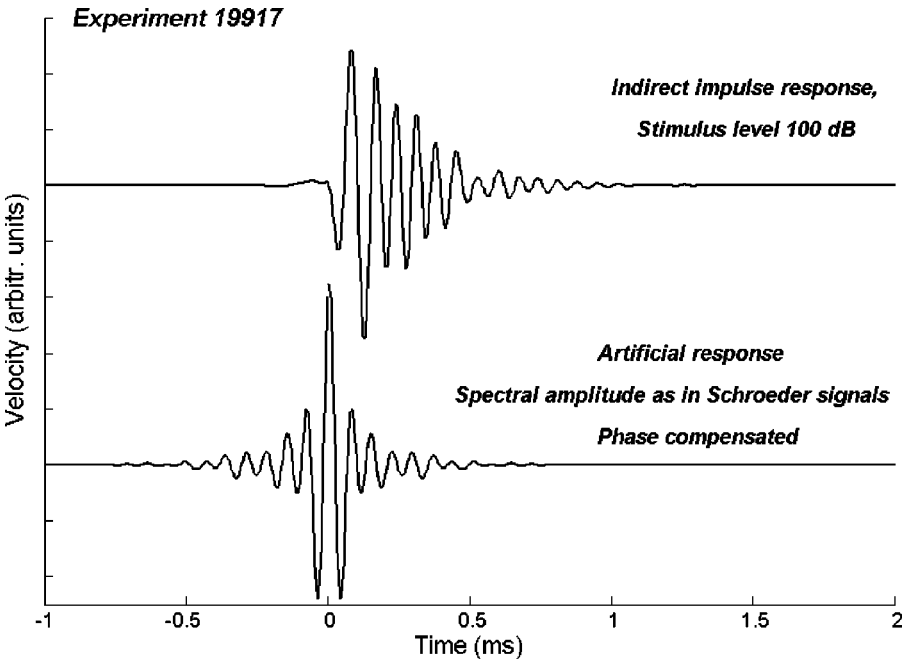


FIG. 10. The ''narrowest response of the cochlea.'' Upper waveform: high-level indirect impulse response (stimulus level 100 dB SPL for the 10–20 kHz octave band). Lower waveform: an artificial response derived from a spectrum of which the amplitude is the product of the spectral amplitude of a Schroeder-type signal and the amplitude of the corresponding (high-level) cochlear transfer function, and the phase is zero.

B. Effects of traveling wave dispersion on response envelopes

The reasoning contained in the preceding paragraphs focuses on the frequency domain and how phase curvature in the stimulus spectrum and the BM transfer function (as derived from the indirect impulse response) combine to influence the shape of the measured response. The origin of phase curvature in the intrinsic BM response appears to lie in *dispersion* of waves traveling in the cochlea. Spatial dispersion can be understood in terms of a (mechanical) model of the cochlea. Such a model generally has the shape of an elongated cavity filled with fluid. The fluid has two outlets, the round window and the oval window (in which the stapes fits). The fluid column is divided into two parts separated by the ''cochlear partition,'' a long structure representing the mechanics of the basilar membrane and the organ of Corti. In the model dispersion arises, first, from inhomogeneity of the mechanics of the cochlear partition (de Boer, 1996, Sec. 4), and, second, from hydromechanical properties of the fluid wave. That wave starts out at the stapes as a ''long'' or ''shallow-channel'' wave (the wavelength is large compared to the cross-sectional diameter of the fluid channel), and gradually turns into a ''short'' or ''deep-channel'' wave (the wavelength is becoming shorter and shorter until it is short compared to the cross-sectional diameter of the fluid channel). A more detailed description of the latter process can be found in Patuzzi (1996) and a theoretical account in de Boer (1996, Sec. 5). The net result of these processes is that low-frequency waves travel faster and arrive earlier at the peak region than high-frequency waves (Shera, 2001). This applies in particular to the basal part of the cochlea. The immediate consequence of this dispersion is found in phase curvature in the frequency domain.

The phase results derived from the indirect impulse data (Sec. III A) agree with response measures reported by others (Sellick *et al.*, 1983; Ruggero *et al.*, 1997) in showing clear evidence of negative phase curvature in the tip region (beginning about half an octave below CF). However, in the low-frequency (tail) region, phase plots—on a linear frequency scale—show little or no curvature, consistent with the notion that these frequencies propagate nondispersively (at a similar rate for all frequencies). With increases in level, linear response growth in low-frequency (''tail'') regions combined with compressive (i.e., less than linear) growth in high-frequency ''tips'' leads to a lowering of the BF and an increase in the relative influence of tail regions on the overall response. Given that components with low frequencies propagate at a nearly constant rate, they arrive essentially unchanged at the measurement location (with the exception of some added, constant delay). Now, consider the changes in response envelopes with level shown for the Schroeder-phase signals and the indirect impulse responses in Figs. 4 and 5. In each experiment and for all three stimulus types, increases in level lead to response envelopes that more and more closely resemble the envelopes of the external (or indirectly modeled) input stimuli. That is, response envelopes become flatter for the two Schroeder-phase signals, and peakier (more impulsive) in the case of the indirect impulse data.

*Non*dispersive propagation of low frequencies is found in the apex of the cochlea where phase by frequency plots for low-frequency components approximate straight lines, indicating propagation delays which are independent of frequency. Consequently, in the apex, the low-to-high frequency glide present in basal regions is not observed (Rhode and Cooper, 1996; Khanna and Hao, 1999; Hemmert *et al.*, 2000; Zinn *et al.*, 2000).

C. Implication for psychophysics

In psychophysical experiments, +Schr complexes are often less effective maskers than corresponding −Schr complexes (Smith *et al.*, 1986; Kohlrausch and Sander, 1995; Carlyon and Datta, 1997a,b; Summers and Leek, 1998; Sum-

mers, 2000; Oxenham and Dau, 2001a,b). Beginning with Smith *et al.* (1986), these masking differences have been interpreted as being linked to the envelope characteristics of BM responses to the two maskers and the idea that a more modulated (peakier) internal response will be produced by an appropriately selected +Schr complex than by a time-reversed (−Schr) version of that stimulus. By these accounts, reduced masking effectiveness for +Schr stimuli is due to probe signal detection during low-amplitude (''quiet'') regions between brief high-amplitude peaks in the internal response to this masker. These differences in the internal response envelopes for +Schr and −Schr complexes have now been verified physiologically, both in chinchillas (Recio and Rhode, 2000) and guinea pigs (present data). In addition, the present data show that the modulation differences in BM responses to +Schr and −Schr stimuli are reduced at high presentation levels where the influence of curvature in the BM phase response is reduced, particularly in the BF region. Greater similarity of response envelopes for the two stimuli at high levels is also consistent with the masking literature: the effectiveness of +Schr and −Schr maskers is very similar at high presentation levels (Summers and Leek, 1998; Summers, 2000).

The masking results cited in the previous paragraph all involved human listeners. To our knowledge, other mammalian species have not been tested psychophysically with Schroeder-phase (or closely related) maskers. However, several species of birds have been tested and show very different results from humans (Leek *et al.*, 2000; Dooling *et al.*, 2001). In three different avian species, +Schr maskers were as effective, or slightly more effective, than −Schr complexes. These results suggest that the negative phase curvature in the BM response observed in mammals may not be present in birds. Although data on BM response in birds are limited, Gummer *et al.* (1987) reported results indicating slight *positive* curvature in BM phase response for pigeons.

The current results demonstrate clear differences in BM response envelopes for +Schr and −Schr complexes composed of identical harmonic components and with extremely similar (very flat) external envelopes. Negative curvature in the intrinsic BM phase response tends to partially cancel the positive phase curvature of +Schr stimuli, producing greater modulation in the BM response than is seen for −Schr stimuli. Responses to the two complexes become increasingly similar as input level increases. The results point out an aspect of cochlear nonlinearity which has previously received little attention. That is, as levels increase, phase curvature in the BF region tends to decrease dramatically. Our results show how the level-dependent changes in response characteristics for the Schroeder-phase stimuli are directly related to and can be determined quantitatively from cochlear nonlinearity.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health (NIDCD-DC-03553 and NIDCD-DC-00141). The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

APPENDIX: UNWRAPPING THE PHASE OF SCHROEDER SIGNALS

Equation (1) leads to phase values that eventually exceed the limits $-\pi$ and $+\pi$ so that every phase value has to be ''unwrapped'' in order to be interpreted. Take, for example, the +Schr stimulus, N_c equal to 100, and n equal to 10. The phase is 0.9π, and for the next value of n it crosses the π boundary and becomes 1.1π, which is equivalent to -0.9π. In the experiments it is the *waveform* of stimulus and response that is recorded, and we wish to derive the relation between these two by using the frequency *spectrum*. For this purpose we apply a Fourier transform to the stimulus and to the recorded waveforms. In the Fourier spectra of the recorded signals all phase values then appear within the limits $-\pi$ and $+\pi$, and the phase functions have to be unwrapped. This unwrapping has to be done in such a way that for the *stimulus* signal the phase function appears (more or less) in the form of Eq. (1), and that a similar form of a monotonically increasing or decreasing function of frequency appears for the *response* phase. Unfortunately, this cannot be done in the ideal way. Figure 11 illustrates the result of the unwrapping procedure for typical signals used in the present work. Consider panel (a). The *thick curve* is the unwrapped phase curve for the +Schr stimulus. For frequencies below 10 kHz this curve follows the form of Eq. (1), but above that frequency the slope of the function suddenly becomes negative. In fact, above that frequency the index n is higher than 50, the quantum steps in the original phase function are larger than π, and the unwrapping routine is unable to follow them and to reconstruct the original phase function accurately. Actually, with every next step of the index n one step of 2π is missed in the unwrapped phase, and the result of unwrapping is a function with negative slope. Note, however, that the *second derivative* of the phase function is not affected (with the exception of one point just prior to where the function reverses direction). Why this is so needs some explanation. To form the second derivative, *three* successive phase values $(\theta_1, \theta_2, \theta_3)$ are needed. The first and the third are added, and two times the central one is subtracted. That in each step of the index n an additional step of 2π is missed does not influence the result for the second derivative. That is, for $x = 0,1,2,3,...,$

$$\theta_1 - 2\theta_2 + \theta_3 = (\theta_1 - 2\pi(x)) - (\theta_2 - 2\pi(x+1)) + (\theta_3 - 2\pi(x+2)).$$

Therefore, the second derivative, i.e., the *curvature* of the phase, is retained. The *thin curve* in panel (a) shows the result of unwrapping the phase of a typical response signal. This curve confirms that the second derivative is not materially affected by the unavoidable errors in the data. Panel (b) of Fig. 11, *thick curve*, shows the corresponding situation for the −Schr stimulus. The *thin line* shows the unwrapped phase for a typical experiment, and, again, the unwrapping

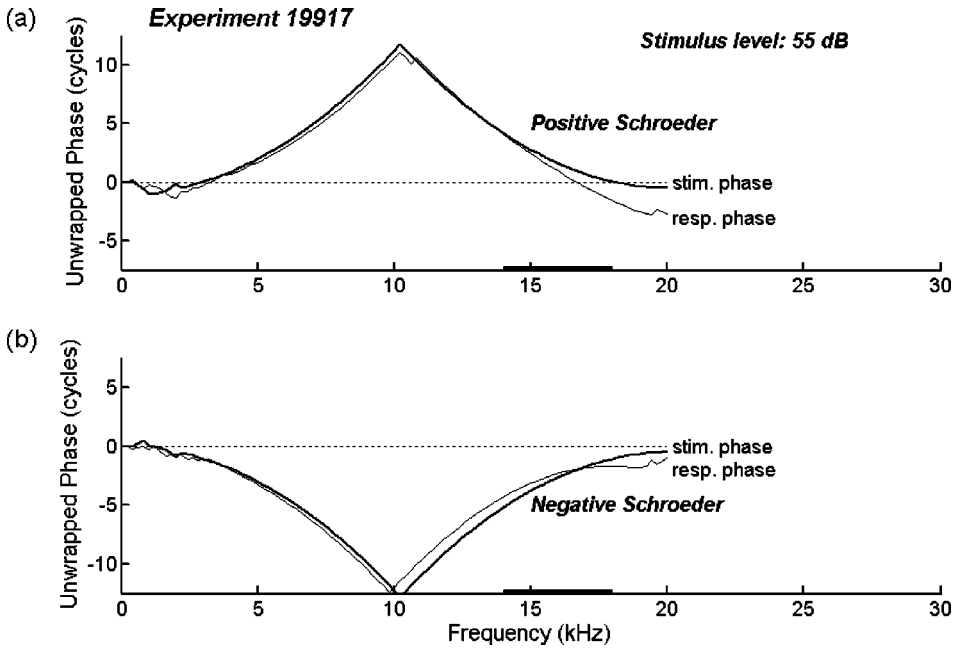


FIG. 11. Unwrapped phase, as a function of frequency. Thick part of the abscissa indicates the region of the response peak (14 to 18 kHz). Thick curve: phase of Schroeder stimulus, thin curve: phase of response. Panel A: positive Schroeder, panel B: negative Schroeder signal. Note how in the upper panel the *phase curvature* is diminished in the peak region, and in the lower panel it is increased in the same region.

routine is accurate enough to reveal the true estimate of the phase curvature.

In Sec. IV it is shown that prediction of responses as if the cochlea is a linear device is pretty accurate. In terms of complex spectra, the spectrum of the stimulus signal is simply multiplied by the cochlear transfer function—which is the spectrum of the cochlear (indirect) impulse response. In this multiplication the phase values are added, and so are the second derivatives. Consider panel (a) of Fig. 11 again. The phase curvature of the spectrum corresponding to the impulse response is negative in the peak region but smaller in magnitude than the positive one corresponding to the stimulus. Hence, the resulting response phase still has a positive curvature but it is smaller than that of the stimulus, and the curvature is smallest in the peak region. Panel (b) shows the corresponding situation for a −Schr signal. In this case the stimulus signal has negative phase curvature and this becomes larger in magnitude in the peak region when the phase of the cochlear transfer function is added.

Figure 12 illustrates the influence of phase curvature in the (linear) frequency domain on the temporal waveform. The figure shows waveforms of signals with a restricted spectrum (see the legend) and phase values that are quadratic functions of frequency. In the waveform plotted in the upper portion of the figure, the phase curvature over the entire frequency range is small and positive [corresponding to the situation in the peak region of panel (a) of Fig. 11)]. The waveform has a negative (downward) glide and a short duration, i.e., it is ''peaky,'' as we call it in the main text. In contrast, phase curvature for the waveform in the lower part of the figure is negative, constant, but four times larger (see the legend). Consequently, the waveform has a long duration. Furthermore, it has a positive (upward) glide. Note the typical ''spindle'' form of the two waveform envelopes; note also

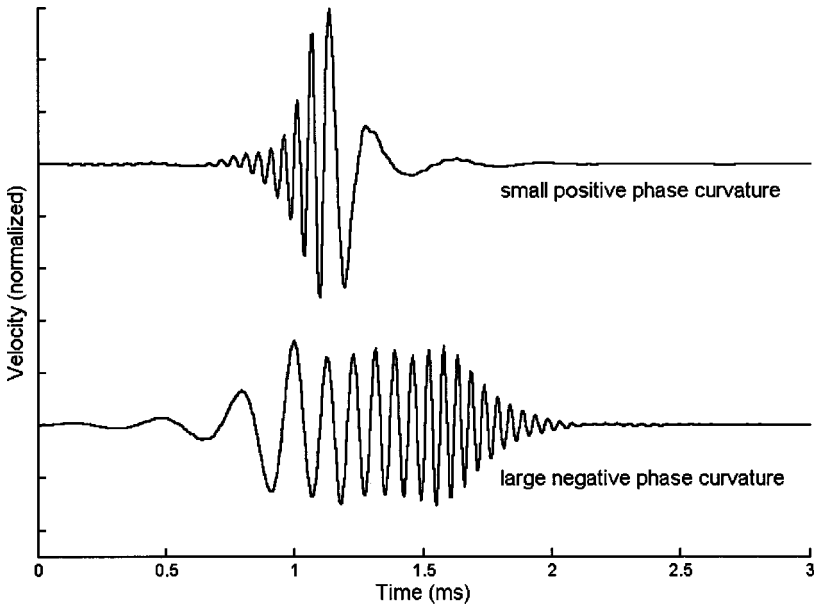


FIG. 12. Waveforms of bandlimited signals of which the phase $\varphi(f)$ of the components is a quadratic function of frequency f. The spectrum is restricted from 2.5 to 22.5 kHz, has constant amplitude, but the outer edges are rounded off. Upper waveform: signal with a positive sign in the phase function. Phase follows (apart from a delay term): $\varphi(f)=+(f/(100f_0))^2$, where f_0 is the sampling frequency (208 kHz). Lower waveform: signal with positive sign in the phase function; the coefficient of the quadratic term is four times larger than in panel (a), the phase follows: $\varphi(f)=-(f/(50f_0))^2$. Note the inversion of the glide in the waveform, and of the direction of the spindle.

that in the upper waveform, the spindle points to the left and, in the lower wave, to the right.

[1]The indirect impulse responses were derived from ''composite-spectrum files'' of which low- and high-frequency parts were acquired separately (de Boer and Nuttall, 1999). The level of the noise then formally refers to the high-frequency part, which was one octave wide.

[2]To avoid considering the presentation of a noncausal signal to the cochlea, a time-shifted (nearly causal) version of the same signal may be substituted. This signal would produce a time-shifted, but otherwise identical response.

Carlyon, R. P., and Datta, A. J. (**1997a**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Carlyon, R. P., and Datta, A. J. (**1997b**). ''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.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). ''Frequency glides in the impulse response of low-frequency auditory-nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.

de Boer, E. (**1996**). ''Mechanics of the cochlea: Modeling efforts,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 258–317.

de Boer, E. (**1997**). ''Connecting frequency selectivity and nonlinearity for models of the cochlea,'' Aud. Neurosci. **3**, 377–388.

de Boer, E., and Nuttall, A. L. (**1997**). ''The mechanical waveform of the basilar membrane. I. Frequency modulations ('glides') in impulse responses and cross-correlation functions,'' J. Acoust. Soc. Am. **101**, 3583–3592.

de Boer, E., and Nuttall, A. L. (**1999**). ''The ''inverse problem'' solved for a three-dimensional model of the cochlea. III. Brushing up the solution method,'' J. Acoust. Soc. Am. **105**, 3410–3420.

de Boer, E., and Nuttall, A. L. (**2000**). ''The mechanical waveform of the basilar membrane. III. Intensity effects,'' J. Acoust. Soc. Am. **107**, 1497–1507.

de Boer, E., and Nuttall, A. L. (**2002**). ''The mechanical waveform of the basilar membrane. IV. Tone and noise stimuli,'' J. Acoust. Soc. Am. **111**, 979–989.

Dooling, R. J., Dent, M. L., Leek, M. R., and Gleich, O. (**2001**). ''Masking by harmonic complexes in birds: Behavioral thresholds and cochlear response,'' Hear. Res. **152**, 159–172.

Geisler, C. D., and Rhode, W. S. (**1982**). ''The phases of basilar membrane motion,'' J. Acoust. Soc. Am. **71**, 1201–1203.

Gummer, A. W., Smolders, J. W. T., and Klinke, R. (**1987**). ''Basilar-membrane motion in the pigeon measured with the Mossbauer technique,'' Hear. Res. **29**, 63–92.

Hemmert, W., Zenner, H.-P., and Gummer, A. W. (**2000**). ''Characteristics of the traveling wave in the low-frequency region of a temporal-bone preparation of the guinea-pig cochlea,'' Hear. Res. **142**, 184–202.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). ''Basilar-membrane measurements and the traveling wave,'' Hear. Res. **22**, 147–153.

Khanna, S. M., and Hao, L. F. (**1999**). ''Nonlinearity in the apical turn of living guinea pig cochlea,'' Hear. Res. **135**, 89–104.

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.

Leek, M. R., Dent, M. L., and Dooling, R. J. (**2000**). ''Masking by harmonic complexes in budgerigars,'' J. Acoust. Soc. Am. **107**, 1737–1744.

Møller, A. R., and Nilsson, H. G. (**1979**). ''Inner ear impulse response and basilar-membrane modeling,'' Acustica **41**, 258–262.

Nuttall, A. L., and Dolan, D. F. (**1996**). ''Steady-state sinusoidal velocity responses of the basilar membrane in the guinea pig,'' J. Acoust. Soc. Am. **99**, 1556–1565.

Oxenham, A. J., and Dau, T. (**2001a**). ''Reconciling frequency selectivity and phase effects in masking,'' J. Acoust. Soc. Am. **110**, 1525–1538.

Oxenham, A. J., and Dau, T. (**2001b**). ''Towards a measure of auditory-filter phase response,'' J. Acoust. Soc. Am. **110**, 3169–3178.

Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.

Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, London).

Recio, A., and Rhode, W. S. (**2000**). ''Basilar-membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.

Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.

Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea,'' Aud. Neurosci. **3**, 101–121.

Rhode, W. S., and Robles, L. (**1974**). ''Evidence from Mossbauer experiments for nonlinear vibration in the cochlea,'' J. Acoust. Soc. Am. **55**, 588–596.

Robles, L., Rhode, W. S., and Geisler, C. D. (**1976**). ''Transient response of the basilar membrane measured in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **59**, 926–939.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Schroeder, M. R. (**1970**). ''Synthesis of low peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar-membrane motion in the guinea pig using the Mossbauer technique,'' J. Acoust. Soc. Am. **97**, 2479–2486.

Sellick, P. M., Yates, G. K., and Patuzzi, R. (**1983**). ''The influence of Mossbauer source size and position on phase and amplitude measurements of the guinea pig basilar membrane,'' Hear. Res. **10**, 101–108.

Shera, C. A. (**2001**). ''Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion,'' J. Acoust. Soc. Am. **109**, 2023–2034.

Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). ''Phase effects in masking related to dispersion in the inner ear,'' J. Acoust. Soc. Am. **80**, 1631–1637.

Summers, V. (**2000**). ''Effects of hearing impairment and presentation level on masking period patterns by Schroeder-phase harmonic complexes,'' J. Acoust. Soc. Am. **108**, 2307–2317.

Summers, V., and Leek, M. R. (**1998**). ''Masking of tones by Schroeder-phase complexes in normally hearing and hearing-impaired listeners,'' Hear. Res. **118**, 139–150.

Zinn, C., Maier, H., Zenner, H.-P., and Gummer, A. W. (**2000**). ''Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the *in vivo* guinea-pig cochlea,'' Hear. Res. **142**, 159–183.

Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey

Mitchell Steinschneider[a)]
Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Rose F. Kennedy Center, Bronx, New York 10461

Yonatan I. Fishman
Department of Neurology, Albert Einstein College of Medicine, Rose F. Kennedy Center, Bronx, New York 10461

Joseph C. Arezzo
Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Rose F. Kennedy Center, Bronx, New York 10461

(Received 13 December 2002; revised 20 April 2003; accepted 25 April 2003)

Voice onset time (VOT) signifies the interval between consonant onset and the start of rhythmic vocal-cord vibrations. Differential perception of consonants such as /d/ and /t/ is categorical in American English, with the boundary generally lying at a VOT of 20–40 ms. This study tests whether previously identified response patterns that differentially reflect VOT are maintained in large-scale population activity within primary auditory cortex (A1) of the awake monkey. Multiunit activity and current source density patterns evoked by the syllables /da/ and /ta/ with variable VOTs are examined. Neural representation is determined by the tonotopic organization. Differential response patterns are restricted to lower best-frequency regions. Response peaks time-locked to both consonant and voicing onsets are observed for syllables with a 40- and 60-ms VOT, whereas syllables with a 0- and 20-ms VOT evoke a single response time-locked only to consonant onset. Duration of aspiration noise is represented in higher best-frequency regions. Representation of VOT and aspiration noise in discrete tonotopic areas of A1 suggest that integration of these phonetic cues occurs in secondary areas of auditory cortex. Findings are consistent with the evolving concept that complex stimuli are encoded by synchronized activity in large-scale neuronal ensembles. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1582449]

PACS numbers: 43.64.Qh, 43.64.Sj, 43.71.Es [WPS]

I. INTRODUCTION

Understanding how information relevant to speech processing is extracted from the neural signal is a cardinal goal of modern auditory neurophysiology. In this study, we examine neural representation of voice onset time (VOT). This important speech parameter is utilized by most of the world languages and signifies the interval between consonant release (onset) and the onset of periodic vocal-cord vibrations (voicing) (Lisker and Abramson, 1964). In American English, a short VOT promotes the perception of the voiced stop consonants /b/, /d/, and /g/, whereas a long VOT facilitates the perception of the unvoiced stop consonants /p/, /t/, and /k/. Differential perception of these phonemes is categorical, with the boundary for discriminating a voiced from an unvoiced stop consonant in syllable-initial position generally lying at a VOT of 20–40 ms.

Various features of speech perception are similar in humans and animals, indicating that phonetic processing relies in part on basic auditory system mechanisms (e.g., Kuhl and Miller, 1975, 1978; Kuhl and Padden, 1982; Sinnott and Adams, 1987; Kluender and Lotto, 1994; Dent *et al.*, 1997; Sinnott and Brown, 1997; Kluender *et al.*, 1998; Ohlemiller *et al.*, 1999; Holt *et al.*, 2001; Le Prell *et al.*, 2001). Animals show categorical perception of voiced versus unvoiced stop consonants with VOT boundary values comparable to those in humans, suggesting that neurophysiological experiments in animals may be directly relevant for investigating mechanisms involved in VOT encoding. Earlier work examining cortical responses to speech sounds in monkey primary auditory cortex (A1) found that syllables with a short VOT often evoked a single transient response at consonant onset, whereas syllables with a long VOT typically evoked two transient responses, one at consonant onset and another at voicing onset (Steinschneider *et al.*, 1994, 1995). These data led to the hypothesis that the perception of voiced from unvoiced stop consonants is partly determined by the difference between these two response patterns.

In this study, we examine the neural representation of stop consonant–vowel syllables with variable VOT in A1, using neuronal population measures in awake, naive old-world monkeys. The goal is to test the hypothesis that previously identified temporal response patterns reflecting the perceptual boundary in humans are maintained in large-scale population activity. Alternatively, these response patterns could be obscured in the overall population activity by other,

[a)]Electronic mail: steinsch@aecom.yu.edu

less specific response components evoked by the syllables, or could be restricted to specific regions of A1. Growing evidence that complex stimuli are encoded by synchronized activity in large-scale neuronal ensembles emphasizes the need to examine population responses sampled across a wide extent of auditory cortex (Creutzfeldt *et al.*, 1980; Gehr *et al.*, 2000; Rotman *et al.*, 2001; Nagarajan *et al.*, 2002; Petersen *et al.*, 2002). Maintenance of temporal response patterns differentially representing VOT in population activity could facilitate the rapid representation of stop consonants that occurs during perception of real-time streams of speech. Studies in awake animals are mandated by the marked effects of anesthesia on auditory cortical and medial geniculate responses to sound (Aitkin and Dunlap, 1968; Mäkelä *et al.*, 1990; Madler *et al.*, 1991; Zurita *et al.*, 1994; Gaese and Ostwald, 2001; Talwar and Gerstein, 2001). Results obtained in naive animals can serve as a benchmark for evaluating physiological changes that occur with learning (e.g., Recanzone *et al.*, 1993; Tremblay *et al.*, 1997; Weinberger, 1997; Ahissar *et al.*, 1998; Merzenich *et al.*, 1999; Blake *et al.*, 2002), a consideration especially relevant for modeling speech-evoked activity in young infants with limited language exposure.

Adequate assessment of speech-evoked activity requires that temporal response patterns be related to organizational features of A1. Thus, the relationship between patterns of activity and the tonotopic organization is examined to assess the degree to which specific response profiles are restricted to discrete best-frequency regions. Laminar-specific responses are also evaluated to determine whether temporal patterns are merely reflections of transmitted subcortical activity or represent newly generated cortical transformations, occur in laminae that project to secondary auditory cortical fields, and are preserved in later polysynaptic activity.

II. METHODS

Six male macaque monkeys (*Macaca fascicularis*) weighing between 2.5 and 3.5 kg were studied following approval by our institutional Animal Care and Use Committee. Animals were housed in our AAALAC-accredited Animal Institute, and their health was monitored daily by the investigators and veterinary staff. Experiments were conducted in accordance with institutional and federal guidelines governing the use of primates. Animals were initially trained to sit comfortably with hands restrained in customized primate chairs. Surgery was then performed using sterile techniques and general anesthesia (sodium pentobarbital). Holes were drilled into the skull to accommodate epidural matrices that allowed access to the brain. Matrices consisted of 18-gauge stainless-steel tubes glued together into a honeycomb form. They were shaped to approximate the contour of the cortical convexity, covered with a protective layer of sterile silastic, and stereotaxically positioned to target A1 at an angle 30 deg from normal to approximate the anterior–posterior tilt of the superior temporal gyrus. This angle permitted electrode penetrations nearly orthogonal to the surface of A1, a requirement for performing one-dimensional current source density (CSD) analysis (e.g., Vaughan and Arezzo, 1988). Matrices and Plexiglas bars permitting painless head fixation were embedded in a mound of dental acrylic secured to the skull with inverted bolts keyed into the bone. Peri- and postoperative anti-inflammatory agents were given to reduce potential discomfort. Recordings began 2 weeks after surgery.

Recordings were performed with multicontact electrodes constructed in our laboratory (Barna *et al.*, 1981). They contained 14 recording contacts arranged in a linear array and evenly spaced at 150-μm intervals ($<$10% error), permitting simultaneous recording across A1 laminae. Contacts were 25-μm stainless-steel wires insulated except at the tip, which were fixed in place within the sharpened distal portion of a 30-gauge tube, and were maintained at an impedance of 0.1–0.4 MΩ at 1 kHz. The reference was an occipital epidural electrode. Headstage preamplification was followed by amplification ($\times$5000) with differential amplifiers (down 3 dB at 3 and 3 kHz). Signals were digitized at a rate of 3400 Hz and averaged by NEUROSCAN software to generate auditory evoked potentials (AEPs). Multiunit activity (MUA) was extracted by high-pass filtering the raw input at 500 Hz (roll-off 24 dB/octave), further amplifying ($\times$8) and full-wave rectifying the derived signal, and computer averaging the resultant activity.

MUA measures the envelope of action potential activity generated by neuronal aggregates, weighted by neuronal location, size, and electrode impedance (see Vaughan and Arezzo, 1988). MUA is similar to cluster activity but has greater response stability than either cluster or single-unit responses (Nelken *et al.*, 1994). We observe sharply differentiated MUA at recording contact spacings of 75 μm (e.g., Schroeder *et al.*, 1990), and other investigators have demonstrated a similar sphere of recording (Brosch *et al.*, 1997). For some recording sessions, data were stored on either an analog (bandpass 0–5 kHz) or a digital tape recorder (sample rate 6 kHz). Due to limitations of the acquisition computer, sampling rates were less than the Nyquist frequency of the low-pass filter setting of the amplifiers. Empirical testing revealed negligible signal distortion, as almost all energy in the neural signals was less than 1 kHz. Samples of off-line data from the digital tape recorder were redigitized at 6 kHz, and resultant MUA had waveshapes and amplitudes nearly identical to those of data sampled at the lower rate (distortion $<$1%). MUA acquired from the digitally taped data was also low-pass filtered below 800 Hz using newer digital filters (96 dB/octave, *RP*2 modules, Tucker Davis Technologies) and then averaged at a sampling rate of 2 kHz to further test the accuracy of the initial measurements. Differences between these and initial measurements were negligible (Fishman *et al.*, 2001). To further validate MUA measures, peristimulus-time-histograms (PSTHs) were constructed from high-pass-filtered (500 Hz) data sampled at 65 kHz with a binwidth of 1 ms (BRAINWARE32, Tucker Davis. Tech.) for all electrode penetrations whose responses were stored on tape. Triggers for spike acquisition were set at 2.5 times the amplitude of the high-frequency background activity. Neuronal cluster responses were acquired from lower lamina 3, the same depth at which MUA was analyzed.

One-dimensional CSD analysis was used to characterize the laminar pattern of net current sources and sinks within

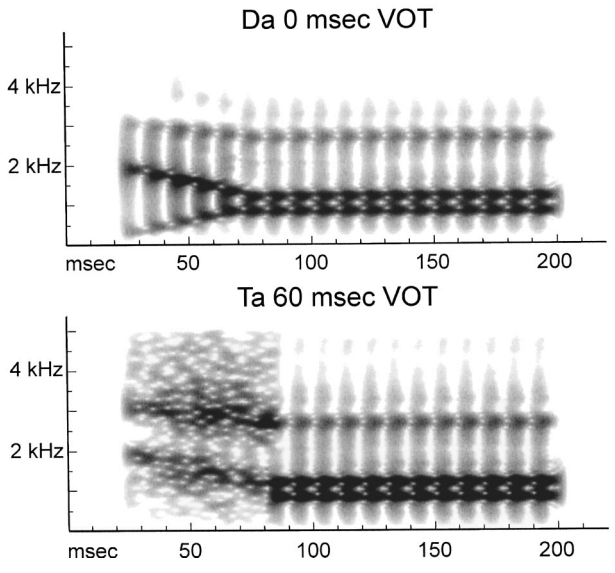


FIG. 1. Sound spectrographs of the synthetic syllables /da/ and /ta/ used in this study. Syllables with a VOT of 0 and 60 ms are shown. Syllables with a VOT of 20 and 40 ms were also presented. See the text for details.

A1. CSD was calculated from AEP laminar profiles using an algorithm that approximated the second spatial derivative of the field potentials across three adjacent depths (Freeman and Nicholson, 1975). Sinks generally index regions of net synaptic depolarization, although they can also represent current return for hyperpolarization occurring at an adjacent site. Sources denote either sites of net current return for nearby depolarizations or locations of hyperpolarization. These possibilities were distinguished by using the concurrently recorded MUA as a measure of net neuronal excitation and inhibition. A sink associated with increased MUA reflects excitatory post-synaptic potentials (EPSPs), while a source associated with MUA reduction indicates hyperpolarization.

Speech sounds were the syllables /da/ and /ta/ with VOTs of 0, 20, 40, and 60 ms. They were initially synthesized at the Haskins Laboratories (New Haven, CT), and later digitized at a sampling frequency of 44.1 kHz and edited. Syllables with a VOT of 0 and 20 ms were reliably perceived by human listeners as /da/, while those with a VOT of 40 and 60 ms were perceived as /ta/. Sound spectrographs of /da/ with a 0-ms VOT and /ta/ with a 60-ms VOT are shown in Fig. 1. Syllables were composed of three formants with steady-state center values of 817 Hz for the first formant ($F1$), 1181 Hz for the second formant ($F2$), and 2632 Hz for the third formant ($F3$). Starting formant frequencies were 200 Hz for $F1$, 1835 Hz for $F2$, and 3439 Hz for $F3$. Formant transitions were 30 ms for $F1$, and 40 ms for $F2$ and $F3$. The fundamental frequency (F_0) was 100 Hz. Syllables were presented at 80 dB SPL. The intensity of aspiration noise at onset was 17 dB less than the sound-pressure peaks in the vowels, and it rose nearly linearly by 5 dB over the 60-ms maximum time span of the noise. Isointensity frequency response functions based on pure-tone responses at 60 dB SPL were used to characterize the frequency sensitivity of the cortical sites and derive an estimate of the best frequency (BF). BF was defined as the frequency that elicited the largest response to tone onset. Pure tones ranged from 0.2 to 12 kHz, with 10-ms linear rise–decay ramps. They were digitally constructed, edited, and delivered at a sampling frequency of 44.1 kHz using DIGIDESIGN TURBOSYNTH and SOUND DESIGNER II software and hardware. All stimuli were 175 ms in duration and were presented once every 658 ms via a dynamic headphone (Sony, MDR-7502) coupled to a 60-cc plastic tube that was placed against the ear contralateral to the recording site. Sound intensity was measured with a Bruel & Kjaer sound-level meter (type 2236) positioned at the opening of the plastic tube attached to the headphone.

Recordings were performed in a sound-attenuated chamber with the animals painlessly restrained. Monkeys maintained a relaxed, but alert state, facilitated by frequent contact and delivery of juice reinforcements. Positioning of the electrode was guided by on-line inspection of AEPs and MUA evoked by 80-dB clicks. Tone bursts and speech sounds were presented when the recording contacts of the linear-array electrode straddled the inversion of early cortical AEP components. Response averages were generated from 50–100 stimulus presentations.

After completion of a recording series, animals were deeply anesthetized with sodium pentobarbital and perfused through the heart with physiological saline and 10% buffered formalin. A1 was physiologically delineated by its typically large amplitude responses and by a BF map that was organized with low BFs located anterolaterally and higher BFs posteromedially (e.g., Merzenich and Brugge, 1973; Morel *et al.*, 1993). Electrode tracks were reconstructed from coronal sections stained with cresyl violet, and A1 was anatomically identified using published criteria (e.g., Morel *et al.*, 1993). Depths of the earliest click-evoked current sinks were used to locate lamina 4 and lower lamina 3 (Steinschneider *et al.*, 1992). This method is compatible with findings of other studies (Müller-Preuss and Mitzdorf, 1984; Metherate and Cruikshank, 1999; Rose and Metherate, 2001; Cruikshank *et al.*, 2002). A later current sink in upper lamina 3 and a concurrent source located more superficially were almost always identified in the recordings. The sink served as an additional marker of laminar depth, and has been repeatedly observed (e.g., Müller-Preuss and Mitzdorf, 1984; Steinschneider *et al.*, 1992, 1994, 1998; Metherate and Cruikshank, 1999; Fishman *et al.*, 2000b; Cruikshank *et al.*, 2002). This physiological procedure for laminar identification was anatomically checked by correlation with measured widths of A1 and its laminae at select electrode penetrations. MUA recorded from infragranular laminae and subjacent white matter, and with onset and peak latencies earlier than intracortical activity, was ascribed to activity in thalamocortical (TC) afferents (Steinschneider *et al.*, 1992, 1994).

Representation of VOT in temporal patterns of population responses was evaluated in several ways. We first examined relative MUA amplitudes within lower lamina 3, segregated according to the BF of the recording sites. Response ratios were derived by dividing the amplitude of the MUA at 10 ms increments after the peak ''on'' response by the maximum amplitude of the initial ''on'' response. Mean peak latencies of the initial ''on'' responses ranged from 12 to 13 ms for the four syllables and were not significantly different

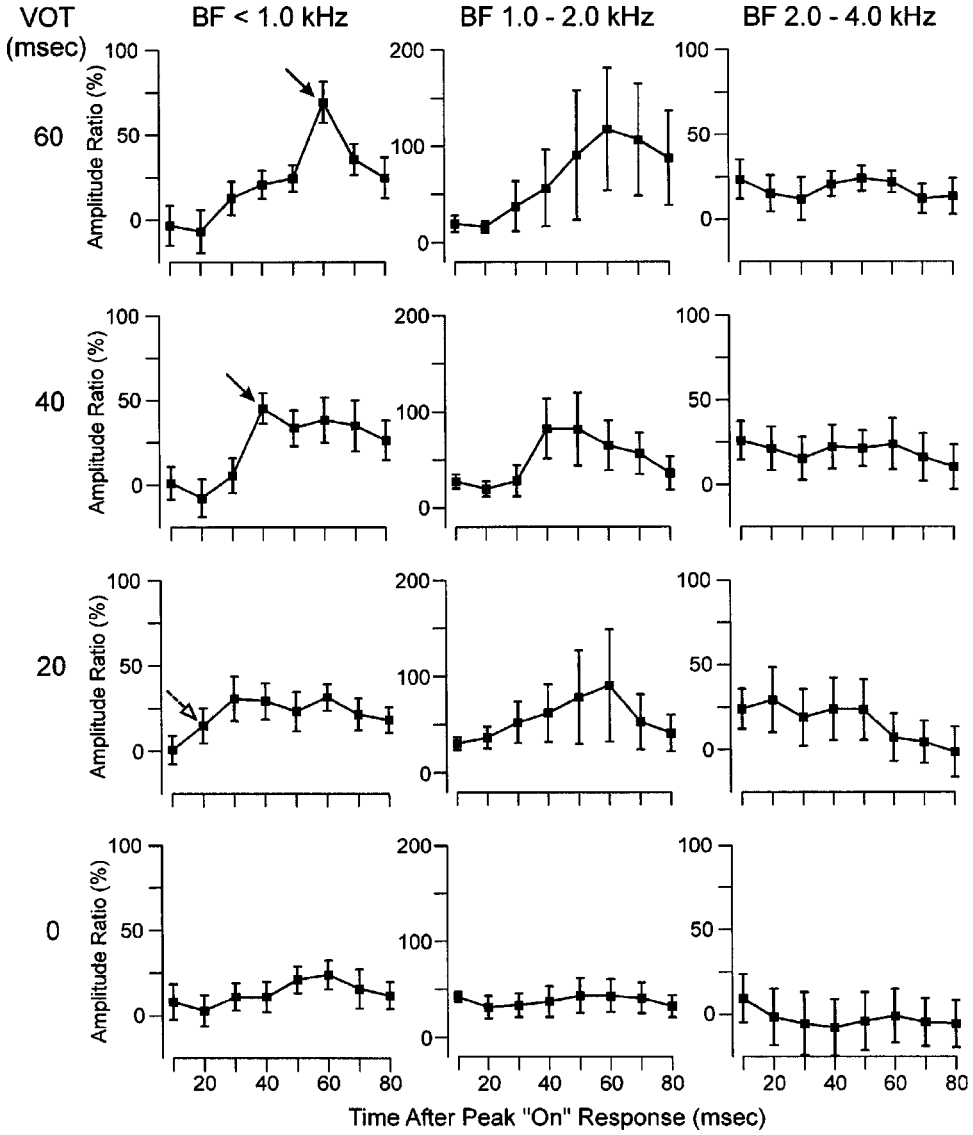


FIG. 2. Response ratios depicting amplitude of lower lamina 3 MUA at 10-ms increments relative to the onset response for each syllable. Ratios are segregated by the BF of the recording site. Syllable VOT is shown at the left. See the text for details.

from each other (ANOVA, $p=0.32$). Data were subjected to a two-way ANOVA, and *post hoc* analysis was performed using a Tukey–Kramer adjustment for multiple comparisons. We also examined the absolute mean amplitude of lower lamina 3 MUA segregated according to BF. Statistical analysis was performed at key time points using one-way ANOVA and *post hoc* Newman–Keuls multiple comparisons test. Significance level for *post hoc* tests was $p<0.05$. Similar analyses were performed for MUA measures of TC fiber activity, PSTH indices, and amplitudes of CSD sinks and sources. Only p values of the ANOVA will be stated.

III. RESULTS

Results are based on 78 electrode penetrations into A1 where lower lamina 3 MUA exhibited a short latency (<25 ms), transient response elicited by syllable onset. MUA had a BF of <1 kHz, 1–2 kHz, 2–4 kHz, and >4 kHz in 33, 15, 12, and 18 penetrations, respectively. Excitation without an ''on'' response was evoked in five additional penetrations, sustained MUA suppression occurred in one penetration, and no responses were obtained in two penetrations.

A. Representation of VOT in A1

Neural representation of VOT is determined by the tonotopic organization of A1. Results are graphically depicted in Fig. 2, which illustrates response ratios obtained in low, middle, and higher BF regions for each stimulus VOT and at each 10-ms incremental time point following the peak of the initial response to stimulus onset. Response patterns differentially reflecting VOT are restricted to lower BF recording sites. At sites with a BF of <1 kHz, response peaks time-locked to voicing onset are observed for syllables with 40- and 60-ms VOTs (solid arrows). This low-frequency region would be predicted to respond to voicing onset, which initiates the increase in $F1$ sound energy. However, the stimulus with a 20-ms VOT fails to evoke a significant peak of activity time-locked to voicing onset (dotted, unfilled arrow). The key statistical result is a significant stimulus×time interaction ($p=0.0001$). *Post hoc* analysis reveals significant increases in activity at voicing onset for the syllables with a VOT of 40 and 60 ms when compared against all preceding time points. This effect is absent in the response evoked by the 20-ms VOT syllable. MUA at locations with BFs between 2–4 kHz does not contain response peaks to voicing

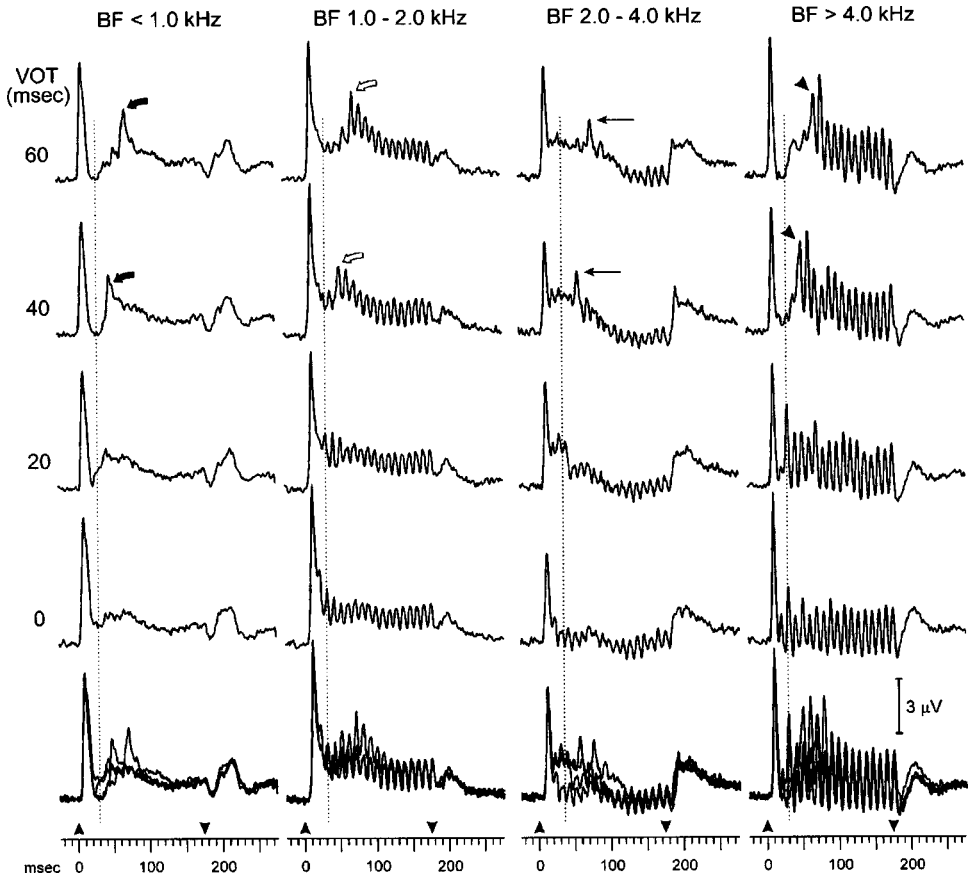


FIG. 3. Averaged absolute amplitude responses in lower lamina 3 evoked by the syllables and segregated by recording site BF. Syllable VOT is shown at the left and superimposed waveforms at the bottom. Arrowheads above the timelines indicate stimulus onset and offset. Vertical dotted lines are aligned with predicted times for responses to voicing onset for the 20-ms VOT syllable. See the text for details.

onset (right-hand column). While the BF region of 1–2 kHz has response peaks evoked by voicing onset for the longer VOT stimuli, there is a marked increase in response variability. Examination of the individual penetrations in this BF range suggests that a differential VOT effect is observed when large-amplitude responses are evoked by lower frequency tones that approximate *F*1.

Similar to results obtained using relative amplitude measures, differential representation of VOT is present in the absolute mean amplitude of synchronously active A1 populations (Fig. 3). Vertical dotted lines are aligned with the predicted location of the peak in the response to voicing onset for the 20-ms VOT stimulus to facilitate waveform comparisons. Overlaid responses to all four syllables are depicted at the figure bottom. Sites with BFs $<$1.0 kHz show peaks in MUA time-locked to both consonant release and voicing onset only in response to syllables with VOTs of 40 and 60 ms (solid curved arrows). MUA elicited by these syllables contains a brief period of suppression that follows the responses to consonant release. This suppression decays in the response to the 60-ms VOT stimulus, and is followed by a plateau of activity upon which the response to voicing onset occurs. The expected response to voicing onset for the 20-ms VOT stimulus occurs at the peak of the MUA suppression, truncating the suppression but restricting the development of a response peak time-locked to voicing onset. ANOVAs performed on response amplitudes at the times of response peaks evoked by voicing onset for the 60- and 40-ms VOT syllables, and at the time of the predicted response to voicing onset for the 20-ms VOT syllable, show significant main effects of stimulus VOT ($p<0.0001$, <0.0029, and <0.0025, respectively). *Post hoc* analyses reveal that peak responses to voicing onset evoked by the 40- and 60-ms VOT syllables are larger than those evoked by the other syllables at the same time points. In contrast, at the time of the predicted response to voicing onset for the 20-ms VOT syllable, activity is similar in amplitude to that elicited by the 0-ms VOT syllable, and both responses are larger than the responses evoked by the prolonged VOT syllables. Somewhat less pronounced changes are observed in the 1.0–2.0-kHz BF region. MUA increases time-locked to voicing onset are still only observed for the prolonged VOT stimuli (unfilled, curved arrows), and ANOVAs reveal main effects of stimulus VOT at the 60-, 40-, and 20-ms VOT time points ($p<0.0001$, 0.0135, and 0.0008, respectively). *Post hoc* tests reveal differences identical to those observed in the $<$1-kHz BF region, with the exception that the response to voicing onset for the 40-ms VOT syllable is not significantly larger at that time point than the activity evoked by the 20-ms VOT stimulus.

A different pattern of activity is seen in the BF region of 2.0–4.0 kHz. Following an initial burst of MUA time-locked to consonant release, there is a sustained plateau of activity that persists throughout the higher frequency aspiration noise. Sustained activity is absent in the response evoked by the 0-ms VOT stimulus, which has no aspiration noise. Another phasic burst of activity that is present for the two longer VOT stimuli (Fig. 3, straight arrows) occurs at the termination of the sustained increase, and is absent in the response to the 20-ms VOT stimulus (vertical, dotted line).

Differences among these latter responses, however, fail to reach statistical significance.

Regions with BFs greater than the spectral content of the syllables (>4.0 kHz) are also responsive to the sounds presented at 80 dB. Response bursts are evoked by consonant onset, and are followed by variable responses occurring during the aspiration noise, and modest increases in activity to voicing onset in the longer VOT stimuli (arrowheads). As for all BF regions greater than 1.0 kHz, periodic activity phase-locked to the 100-Hz F_0 is present. These modulations are in-phase across the neuronal populations of the three higher BF regions.

Similar response patterns reflecting VOT are observed in PSTHs derived from lower lamina 3 cell cluster activity (data not shown). Responses are drawn from BF regions where MUA contained components time-locked to consonant release and voicing onset: BF was less than 1 kHz in 18 penetrations, and 1–2 kHz in 10 others. PSTHs contain response components time-locked to both consonant and voicing onset for the 40- and 60-ms VOT stimuli, while the 0- and 20-ms VOT syllables evoke a response only to consonant release followed by a plateau of sustained activity. ANOVAs reveal significant main effects of stimulus VOT at the 60-, 40-, and 20-ms VOT time points ($p<0.0128$, 0.05, and 0.0006, respectively). *Post hoc* analyses indicate that responses to the 40- and 60-ms VOT syllables are larger at the time of their voicing onsets than activity evoked by the shorter VOT syllables at the same time points. The response to the 20-ms VOT syllable at the expected peak to voicing onset is not different from that evoked by the 0-ms VOT sound, and both are larger than responses to the 40- and 60-ms VOT syllables.

B. Thalamocortical fiber (TC) representation of VOT

Cortical responses could simply reflect transmission of a preexisting pattern present in the activity of TC fibers, or alternatively, they could represent a transformation of the input activity. This issue was addressed by examining MUA from infragranular laminae and subjacent white matter recorded simultaneously with that in lower lamina 3. MUA was ascribed to activity in TC afferents based on its depth and earlier response latency. TC fiber MUA was identified in 22, 7, and 7 electrode penetrations that entered low, middle, and higher BF regions of A1, respectively. The average separation between MUA recorded from lower lamina 3 and deeper activity in TC afferents was 0.9, 0.8, and 1.2 mm in the low, middle, and high BF groups, respectively. Peak latencies of the initial ''on'' responses of TC fibers were 3, 2, and 2 ms earlier than those of lower lamina 3 MUA for the three BF groups, respectively.

TC fiber MUA recorded from low BF regions of A1 also exhibits accentuated response peaks evoked by voicing onset in the 40- and 60-ms VOT stimuli (Fig. 4, solid arrows). Responses evoked by the 0- and 20-ms VOT stimuli are very similar in morphology. The absent peak in the response to the 20-ms VOT stimulus is indicated by the unfilled arrow. While this finding indicates that the cortical MUA is partly based on subcortical response patterns, intracortical processes appear to amplify differences between responses

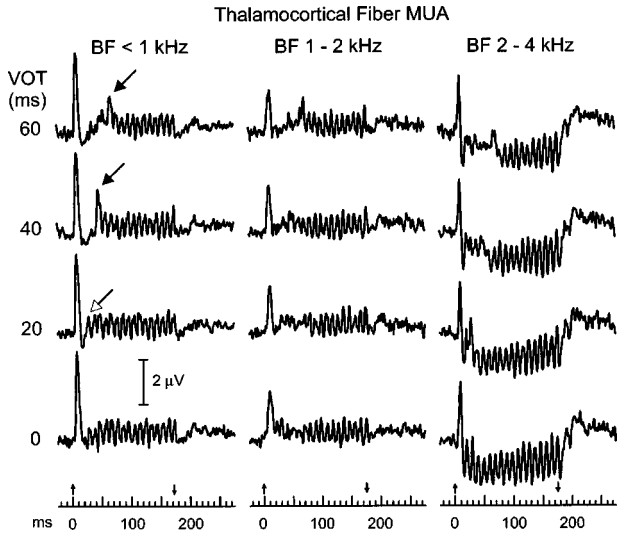


FIG. 4. TC fiber MUA recorded from infragranular laminae and white matter immediately beneath A1. Response peaks to voicing onset are evoked by the prolonged VOT stimuli but not by the syllable with the 20-ms VOT in the low-BF region (arrows). Low amplitude response peaks to voicing onset, especially to the 60-ms VOT syllables, are also observed in the TC fiber MUA recorded in higher BF regions. Additional speech features are represented in TC fiber responses. In all BF regions, the 100-Hz F_0 is represented by phase-locked responses. This pattern is absent from the cortical MUA within low BF regions (Fig. 3), paralleling findings using click train and complex tone stimuli (Steinschneider *et al.*, 1998: Fishman *et al.*, 2000a). $F1$ duration is tracked by sustained increases in MUA within BF regions less than 1 kHz and by sustained suppression in the highest BF region, patterns also observed in the cortical MUA (Fig. 3).

evoked by short and long VOT syllables. The TC fiber response peak evoked by voicing onset in the 60-ms VOT syllable fails to reach statistical significance when compared against responses evoked by the other syllables at the same time point ($p<0.15$). While this failure could be due to the decreased sample size relative to the lower lamina 3 MUA ($N=22$ vs 33), the same analysis performed on lower lamina 3 MUA from these 22 penetrations demonstrates a significant response increase at voicing onset ($p<0.0001$). Similarly, while TC fiber response amplitudes at the time of voicing onset for the 40-ms VOT syllable are significantly different from each other ($p<0.03$), no *post hoc* test is significant. In contrast, *post hoc* analysis performed on the more limited cortical sample does demonstrate a larger response to voicing onset for the 40-ms VOT syllable when compared against activity evoked by the other syllables at the same time point (associated ANOVA, $p=0.0007$).

CSD profiles recorded in lower lamina 3 partly reflect the initial synaptic activity in A1 evoked by its thalamic input, and display a differential pattern of VOT representation that is also less pronounced than the MUA in lower lamina 3 (Fig. 5). Following large amplitude sinks evoked by consonant release, low amplitude sinks are generated by the syllables with a 40- and 60-ms VOT in lower BF regions (left-hand column, solid arrows). Their low amplitude may be partly based on net current summation with current sources evoked by other concurrent synaptic events. Despite their small size, activity evoked by voicing onset in the four syllables is significantly different ($p<0.032$ and $p<0.0001$ for the 60- and 40-ms VOT stimuli, respectively), and *post*

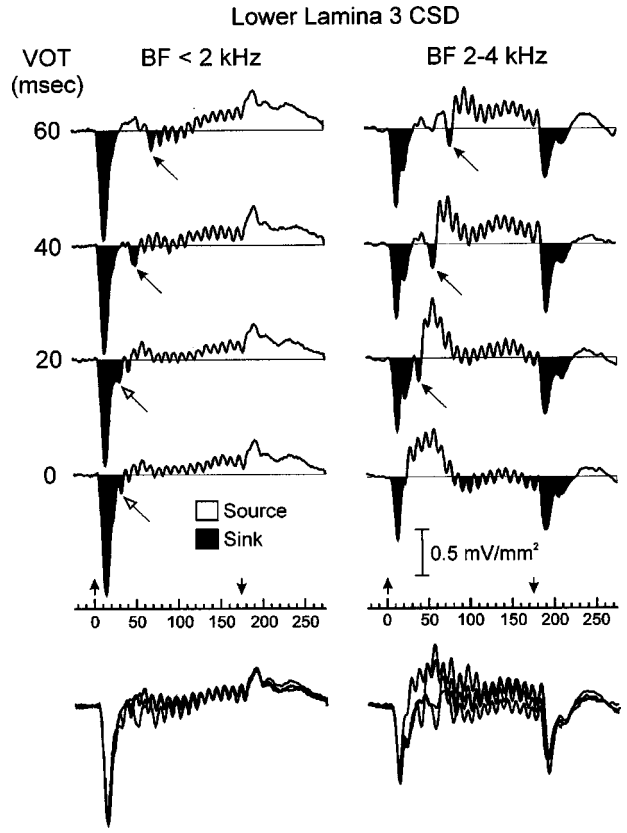


FIG. 5. Averaged CSD profiles recorded from lower lamina 3 segregated by BF of the recording sites. Superimposed waveforms are shown at the bottom. See the text for details.

hoc analyses indicate that the sinks evoked by the 40- and 60-ms VOT syllables are significantly larger than the activity evoked by the other two syllables at the same time points. While activity at the time point for the expected peak in the response to voicing onset for the 20-ms VOT sound is different among the syllables ($p<0.005$), the sinks evoked by the 0- and 20-ms VOT syllables are equivalent (unfilled arrows), and are greater than the responses to the longer duration VOT sounds. A different CSD pattern is observed in the higher BF region (right-hand column). Here, sinks occurring after the termination of the aspiration noise are seen for the syllables with 20-, 40-, and 60-ms VOTs (arrows), followed by sources and sinks evoked by syllable offsets.

Cortical accentuation of responses reflecting VOT is exemplified by the laminar profile of MUA simultaneously recorded across middle and lower laminae during an electrode penetration into a lower BF area (Fig. 6). Latencies of MUA peaks at depths A–F are more than 2 ms later than those recorded at depths G–I. The very short onset latency of the deeper responses (~5 ms) is indicative of their TC fiber origin. Tone-evoked responses are sharply tuned and are maximal to frequencies below 1 kHz at all sites except depth H, which has a broad tuning curve spanning 0.2 to 8 kHz. MUA at the three deeper sites (G–I) contains prominent bursts evoked by syllable onset followed by activity phase-locked to the F_0 and variable increases time-locked to voicing onset. In contrast, MUA at the more superficial depths (A–F) contains prominent responses to both syllable and voicing onsets for the longer VOT stimuli (solid arrows) and an absence of

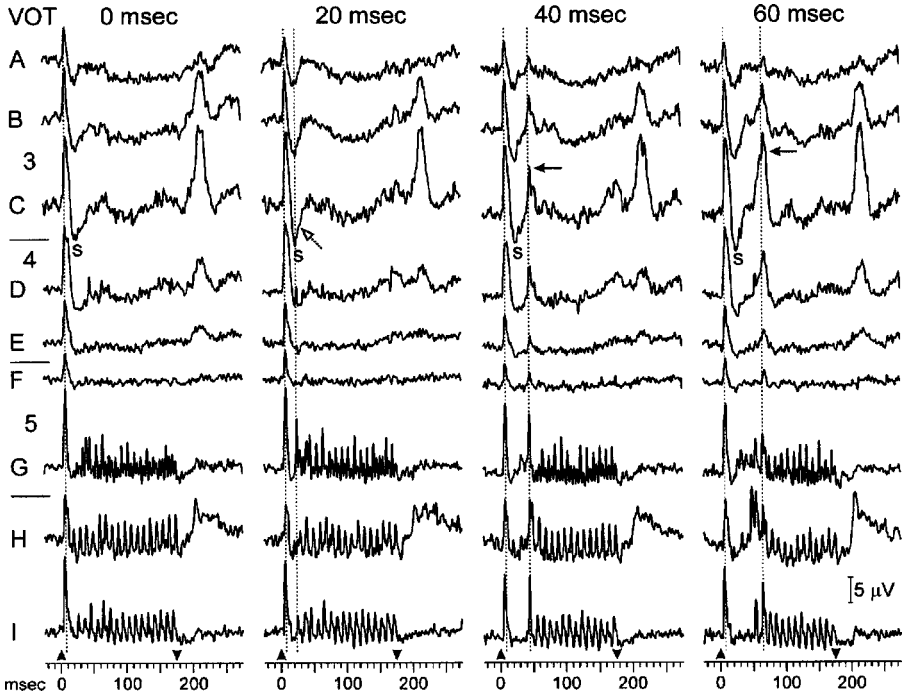


FIG. 6. Laminar profile of MUA simultaneously recorded at 150-μm intervals during an electrode penetration into a low BF site in A1. Each recording depth is labeled at the far left, where approximate laminar boundaries are also shown. Note the difference in response patterns from the cortical MUA recorded at depths A through F, and those from TC fibers recorded at depths G through I. Dotted lines facilitate comparisons across responses at consonant and voicing onsets, and help illustrate the greater than 2-ms latency shift in the TC fiber responses and those from cortex. Following the response burst to consonant onset, there is a period of MUA suppression denoted by the symbol S. The predicted response to voicing onset for the 20-ms VOT syllable occurs during this suppression (dotted arrow), and the response is similar to that evoked by the 0-ms VOT stimulus. In contrast, both consonant and voicing onset elicit phasic responses for the two longer VOT stimuli (solid arrows).

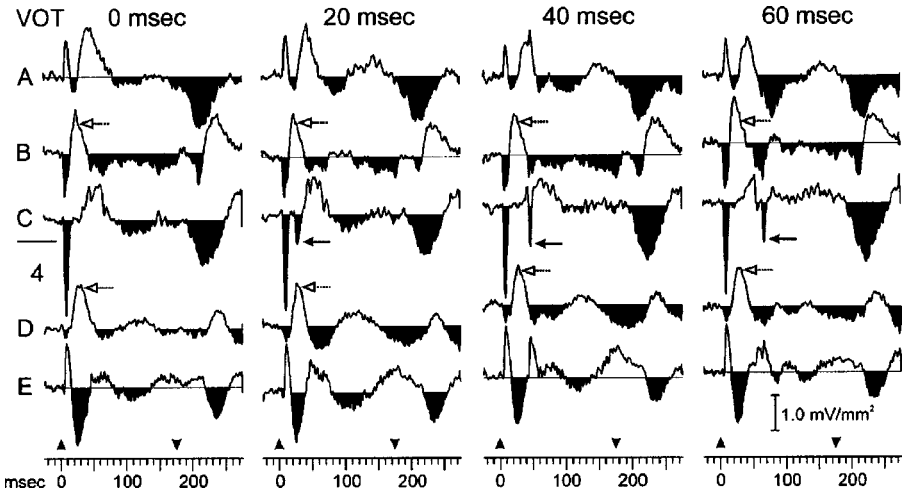


FIG. 7. Laminar CSD profile recorded simultaneously with the MUA shown in Fig. 6. Depths shown at the left correspond to the same depths in the previous figure. Note that responses to voicing onset at depth C in the CSD are present to syllables with VOTs of 20, 40, and 60 ms (solid arrows), while the cortical MUA shown in Fig. 6 has responses to voicing onset for only the two more prolonged VOT stimuli. MUA suppression in the previous figure is coincident with current sources at depths B and D (unfilled arrows).

phase-locked activity to the F_0. The location of the predicted response to voicing onset for the 20-ms VOT syllable is shown by the unfilled dotted arrow. At the time when the response to voicing onset should occur, MUA suppression below baseline levels (denoted by S) is present throughout depths A–F. This suggests that MUA patterns differentially reflecting VOT are accentuated by transient neuronal suppression.

Simultaneously recorded CSD supports this conclusion (Fig. 7). Sinks located at depths B–D mark the initial cortical depolarizations. At the center of the initial depolarization (depth C), additional sinks evoked by voicing onset are observed for the syllables with VOTs of 20, 40, and 60 ms (solid arrows). This pattern is less restrictive than the simultaneously recorded MUA (Fig. 6), which fails to reveal a cortical response to voicing onset for the 20-ms VOT syllable. However, the CSD does resemble the pattern of TC fiber MUA recorded at depth G, suggesting that the CSD represents activity derived from TC fiber input. This change in temporal response patterns between TC fiber MUA and subsequent synaptic activity, and the intracortical MUA, supports the conclusion that intracortical processes sharpen responses reflecting VOT. Sources concurrent with MUA suppression (Fig. 7, unfilled dotted arrows) likely reflect currents induced by inhibitory events.

C. Interlaminar processing of VOT

Early cortical response patterns reflecting syllable VOT are maintained in subsequent interlaminar processing within regions with BFs<2 kHz (Fig. 8). The left-hand column depicts the averaged CSD recorded in upper lamina 3, while the right-hand column illustrates CSD recorded from even more superficial laminae. The averaged CSD recorded from these upper lamina 3 sites consists of an initial current sink evoked by consonant release with onset and peak latencies 5–6 ms later than those in lower lamina 3. Sinks evoked by voicing onset (arrows) show a marked amplitude reduction

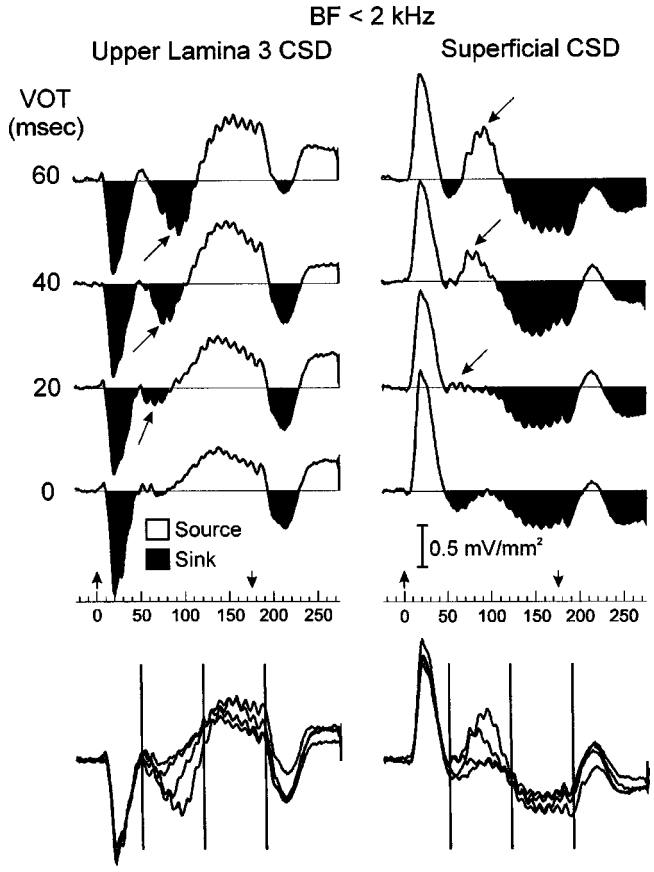


FIG. 8. Averaged upper lamina 3 and more superficial CSD recorded from penetrations into A1 areas with BFs less than 2 kHz. Average separation between the earlier lower lamina 3 sink and the more superficial and later sink in upper lamina 3 is 380 μm, while the even more superficial activity was recorded an average of 333 μm above the upper lamina 3 CSD. Superficial sources are coincident with the upper lamina 3 sinks, consistent with a current dipole within apical dendrites of pyramidal cells. Response evoked by voicing onset to the longer VOT syllables is greater than that elicited by the 20-ms VOT syllable (arrows). Effects are highlighted by the superimposed waveforms shown at the bottom of the figure, which also illustrates boundary demarcations used to quantify the data.

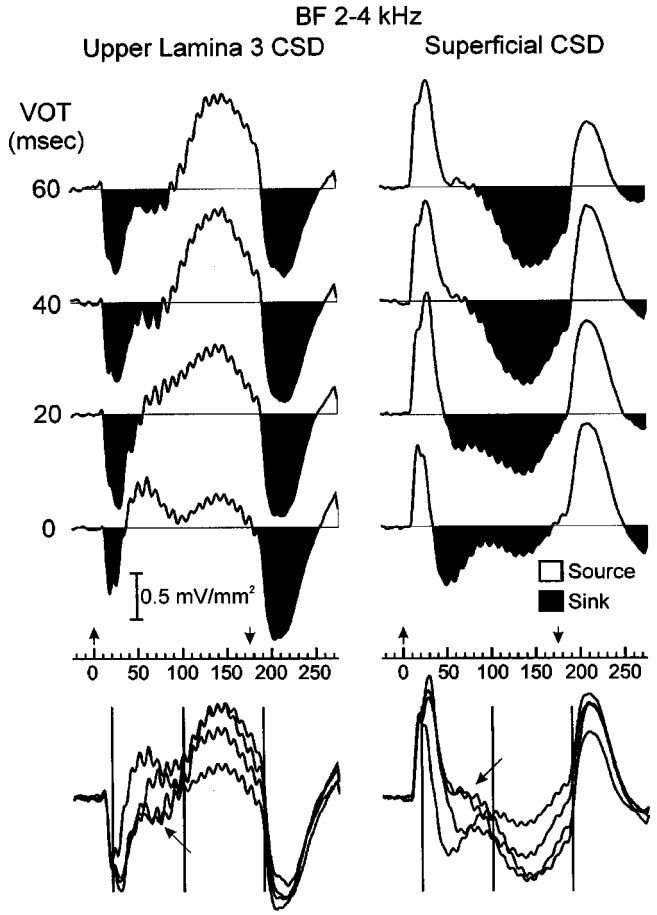


FIG. 9. Averaged upper lamina 3 and more superficial CSD recorded from penetrations into A1 areas with BFs of 2–4 kHz. Separation between the lower lamina 3 and upper lamina 3 recording sites averages 355 μm, while separation between the latter depth and the superficial CSD is 368 μm. Arrows overlying the superimposed waveforms at the bottom of the figure highlight the extended sources and sinks evoked by the 40- and 60-ms VOT syllables. Vertical lines denote boundaries for the areas used to quantify the data.

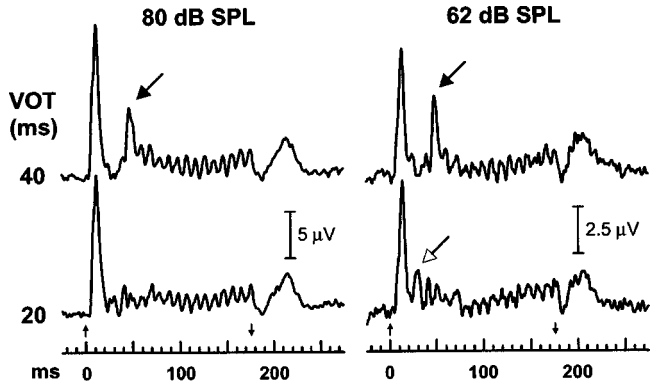


FIG. 10. Averaged lower lamina 3 MUA evoked by syllables presented at the standard 80-dB SPL intensity and at 62 dB SPL. Activity represents the mean of the responses recorded from six A1 sites, three located in areas with BFs <1 kHz and three with BFs between 1 and 2 kHz. Solid arrows mark the accentuated responses evoked by voicing onset of the 40-ms VOT syllable. There is an increase in the ratio of responses evoked by voicing onset relative to the activity evoked by consonant release for the less intense syllables, including that elicited by the 20-ms VOT stimulus (unfilled arrow).

when elicited by the 20-ms VOT syllable. More superficial CSD patterns mirror these findings (arrows). This dipole configuration is consistent with synaptic activation of pyramidal cell apical dendrites in lamina 3 with passive current return occurring in more superficial laminae. Effects are quantified by examining the area of sources and sinks evoked by the various stimuli. Two portions of the waveforms are examined. The first encompasses the time interval from 50 to 120 ms, and the second incorporates the interval from 120 to 190 ms. Interval boundaries correspond to points of common divergence in upper lamina 3 between the ends of the initial sink, the period ending the second sink evoked by voicing onset and start of the following source, and the CSD response elicited by stimulus offset (Fig. 8, drop lines in superimposed waveforms). In the 50–120-ms time interval, responses evoked by the syllables are significantly different from one another in both upper lamina 3 and the more superficial CSD (both $p<0.0001$). In both cases, the syllable with the 60-ms VOT produces a larger response than the syllables with a VOT of 0 and 20 ms, while the response to the 40-ms VOT syllable is greater than that to the 0-ms VOT sound. Additionally, the source in the superficial CSD evoked by the 40-ms VOT sound is larger than that evoked by the 20-ms VOT syllable. These effects are best illustrated by the superimposed waveforms shown at the bottom of the figure.

Aspiration noise modulates interlaminar processing in the 2–4-kHz BF region (Fig. 9). The prolonged VOT syllables elicit extended sinks in upper lamina 3 and sources in the more superficial CSD (arrows overlying superimposed waveforms at figure bottom). Sources and sinks evoked by the 20-ms VOT sound are intermediate in amplitude between those elicited by the /ta/ stimuli and those generated by the 0-ms VOT syllable. Areas of sources and sinks were computed from the intervals of 20 to 100 ms, and 100 to 190 ms, boundaries representing common points of diversion in the waveforms. The magnitude of the sink from 20 to 100 ms in upper lamina 3 is the only difference reaching statistical significance ($p=0.02$), with *post hoc* analysis indicating that the responses to the longer VOT syllables are larger than the activity evoked by the 0-ms VOT syllable.

D. Effects of stimulus intensity

Comparisons between syllable-evoked activity elicited by stimuli presented at 80 and 62 dB SPL were made at a limited number of A1 sites. Three sites had BFs<1 kHz, while 3 had BFs between 1 and 2 kHz. The averaged MUA responses elicited by the syllables with a 20- and 40-ms VOT are shown in Fig. 10. Accentuated responses evoked by voicing onset of the more prolonged 40-ms VOT stimulus persist despite attenuation of syllable intensity (filled arrows) and a twofold decrease in overall response amplitudes. There is for the softer stimuli, however, an increase in the ratio of responses evoked by voicing onset relative to that elicited by consonant onset that occurs for both the 40-ms and the 20-ms VOT stimuli (unfilled arrow).

IV. DISCUSSION

A. Summary of findings and relationship to psychoacoustic studies

The present study demonstrates that features of speech important for discriminating voiced (e.g., /d/) from unvoiced (e.g., /t/) stop consonants are represented by synchronized responses in A1 neuronal populations. VOT is the principal determinant of this phonetic perception (Lisker and Abramson, 1964; Faulkner and Rosen, 1999). Syllables which are generally perceived as /ta/ evoke statistically reliable responses to both consonant release and voicing onset in A1, whereas syllables which are usually perceived as /da/ elicit statistically significant activity only to consonant release. Syllables generating a ''double on'' response pattern could be rapidly placed into the category of unvoiced stop consonants, whereas those eliciting a ''single on'' response could be identified as a voiced consonant. Discrimination between syllables that straddle these response patterns would require only that the brain differentiate between ''single on'' and ''double on'' activity profiles, whereas more subtle timing discriminations would be required for differentiating syllables residing on the same side of the perceptual boundary (Pisoni and Lazarus, 1974; Carney *et al.*, 1977; Pisoni *et al.*, 1982; Kewley-Port *et al.*, 1988). This scheme is consistent with the idea that phonetic encoding is partly based on a ''warped'' representation of acoustical properties, such that some acoustic differences located along a specific portion of a physical continuum are readily distinguished from one another, whereas similar magnitude differences located along a different portion of that continuum are difficult to discriminate (Stevens, 1981).

Three main categories of VOT relationships occur in speech (Lisker and Abramson, 1964). Voicing either begins before (lead), after (lag), or near the time of consonant release, and there is a nonoverlapping trimodal distribution of VOTs for almost all the world's languages. Voicing lead falls in the range of -125 to -75 ms, voicing near the time of consonant release at 0 to $+25$ ms, and voicing following release at $+60$ to $+100$ ms. Averages for these three categories are -100, $+10$, and $+70$ ms, respectively. An A1 physiological boundary of between $+20$ and $+40$ ms falls well within the range of the latter two VOT categories. This temporal processing scheme is not negated by changes in speaking rate modulating VOT (Summerfield, 1981; Utman, 1998; Allen and Miller, 1999; Boucher, 2002). While faster rates shorten VOTs of unvoiced, aspirated stops, values remain greater than 40 ms. Temporal patterns alone, however, cannot distinguish between voicing lead and voicing lag categories, as both would yield ''double on'' responses. Representation of spectral cues such as the presence of low-frequency voicing energy prior to consonant release, or high-frequency aspiration noise after consonant release, could facilitate this discrimination.

The $+20$- to $+40$-ms VOT interval likely represents a natural psychoacoustical boundary in mammalian hearing. Developmentally, young infants with little language exposure can discriminate a VOT contrast of $+20$ with $+40$ ms, even when this contrast is not phonetically relevant for the native language of the child (Eimas *et al.*, 1971; Lasky *et al.*, 1975; Eilers *et al.*, 1979; Jusczyk *et al.*, 1989). An interval of 20–40 ms is generally required to perceive the temporal order of two acoustic events (Hirsh, 1959; Kewley-Port *et al.*, 1988; Jusczyk *et al.*, 1989), supporting the notion that VOT perception is partly based on the ability to determine whether consonant release and voicing onset are sequential in time (Pisoni, 1977). Many animal species respond to VOT discriminations in a categorical manner with perceptual boundaries of $+20$ to $+40$ ms (e.g., Kuhl and Miller, 1975, 1978; Kuhl, 1981; Kuhl and Padden, 1982; Sinnott and Adams, 1987; Dooling *et al.*, 1989). ''Single on'' vs ''double on'' temporal response patterns could serve as neurophysiological cues for all these perceptual phenomena.

Temporal response patterns do not directly account for VOT boundary shifts that occur with changes in consonant place of articulation. Boundaries are shortest for the distinction between /b/ and /p/ ($\sim$20 ms), longest for /g/ and /k/ ($\sim$40 ms), and intermediate for /d/ and /t/ ($\sim$30 ms) (Lisker and Abramson, 1964). Boundary shifts similar to those in humans are observed in animals, suggesting that the phenomenon is not based on language-specific processes (Kuhl and Miller, 1978). $F1$ frequency at voicing onset is the key acoustic feature responsible for this boundary shift (Lisker, 1975; Summerfield and Haggard, 1977; Soli, 1983). It is highest for labial stops, lowest for velar stops, and intermediate for alveolar stops. As $F1$ frequency is lowered, a longer VOT is required for perceiving an unvoiced stop consonant. Greater frequency differences between two consecutive tones increase the likelihood that they will be perceived as occurring simultaneously (Parker, 1988). Japanese quail are more likely to label stop consonants as unvoiced when they incorporate a high $F1$ frequency (Kluender, 1991; Kluender and Lotto, 1994). These studies implicate auditory system mechanisms and stress the importance of $F1$ spectral cues interacting with VOT in modulating discrimination of voiced from unvoiced consonants. Preliminary data in both humans and monkeys suggest that the capacity of A1 to generate one vs two onset responses is determined by the frequency disparity between $F1$ and higher frequency components (Steinschneider *et al.*, 2000), supporting the capability of these temporal response patterns to account for the perceptual VOT boundary shifts that occur with changes in consonant place of articulation.

Aspiration noise is an additional cue important for discrimination of voiced from unvoiced stop consonants that is represented in patterns of A1 activity (Repp, 1979; Sinnott and Adams, 1987; Kluender *et al.*, 1995; Lotto and Kluender, 2002). The likelihood of perceiving an unvoiced stop parallels increases in aspiration noise intensity. This acoustic cue is represented by sustained activity extending throughout its duration in higher BF regions.

Syllable intensity is an important variable that was not thoroughly examined in the present study. Perception of unvoiced stop consonants diminishes with decreases in syllable intensity (Kluender *et al.*, 1995: Lotto *et al.*, 2002). This effect cannot be satisfactorily predicted on the basis of whether low BF areas in A1 produce a ''single'' or ''double on'' neural response. Our limited data obtained at lower stimulus

intensities suggest that as intensity is lowered there is an increased relative response to voicing onset. This would suggest more unvoiced stop perceptions at softer intensities (i.e., a greater number of ''double on'' neural responses), and runs contrary to the perceptual data. A different interpretation offered by a reviewer suggests that as response amplitude evoked by voicing onset increases relative to that evoked by consonant onset, a bias towards a percept of a voiced stop consonant would ensue. This mechanism could be especially important at stimulus intensities where the aspiration noise is near or below threshold. In either case, a key component of the intensity effect is the relative loudness of the aspiration noise (Repp, 1979; Kluender *et al.*, 1995; Lotto *et al.*, 2002). One can predict that as intensity decreases, responses to this lower amplitude component in higher BF regions of A1 would also diminish, thereby decreasing aspiration noise as a potential cue for unvoiced stops and shifting VOT perception towards voiced stops. This physiological result could in turn counterbalance any opposite effect on perception based on activity within lower BF regions. Given that VOT and aspiration noise are represented in different tonotopic areas of A1, it follows that integration of these cues does not occur at this level of the auditory system. Secondary areas of auditory cortex, however, do integrate activity from multiple regions of A1 (Rauschecker *et al.*, 1995), implicating these downstream cortical fields as possible sites that will ultimately determine VOT perception.

B. Relationship to other auditory physiological studies

Many response features similar to those seen in A1 are also observed in auditory-nerve fibers and cells of the auditory brainstem, indicating that the auditory periphery initiates a physiological representation of VOT that is maintained through the auditory pathways. In low BF auditory-nerve fibers, syllables evoke a transient response to consonant release, a period of suppression, and a response to voicing onset for syllables with VOTs greater than 20 ms (Sinex and McDonald, 1988). Response increases evoked by voicing onset are abrupt for syllables with VOTs of 30–40 ms, and more gradual for longer VOT values. Activity in low BF regions of A1 parallel these peripheral patterns (see Figs. 2 and 3). Within the inferior colliculus, responses evoked by consonant release and voicing onset usually merge into a single burst at a VOT less than 20 ms (Chen *et al.*, 1996), though response patterns are also modulated by interaural time differences for binaurally presented syllables (Chen and Sinex, 1999).

We indirectly examine activity in medial geniculate by investigating TC fiber response patterns, and find population responses that differentially reflect VOT. These response patterns, while consistent with continued transmission of VOT-related information from inferior colliculus, are not as robust as those seen in A1. Findings parallel our earlier observations on the transformation of activity between TC fibers and neuronal populations in A1 (Steinschneider *et al.*, 1994), and indicate that one effect of A1 processing is to accentuate responses to lower rate acoustical transients. This accentuation has been previously observed (e.g., Creutzfeldt *et al.*, 1980; Miller *et al.*, 2001), and may facilitate the representation of VOT.

Present findings are compatible with those of other A1 investigations. Spatio-temporal profiles of activity in cat A1 evoked by the onsets of /be/ and /pe/ are similar, whereas a second burst of activity is evoked by the 60-ms VOT of /pe/ in low BF regions (Schreiner, 1998). A boundary of 15–20 ms is seen in the ability to generate a ''double on'' response to /pa/ in epidural recordings overlying guinea pig auditory cortex (McGee *et al.*, 1996). Reliable ''double on'' profiles are observed for syllables with VOTs of 30 ms or longer at stimulus levels of 45–65 dB SPL (Eggermont, 1995a). For syllables presented at 45–75 dB, the ratio of the response evoked by voicing relative to that evoked by consonant onset markedly increases when the VOT is greater than 25 ms (Eggermont, 1995b). These temporal response patterns are not restricted to syllables varying in their VOT, and similar activity occurs for gap-in-noise stimuli (Eggermont, 1995b), repetitive frequency-modulated tones (Lu *et al.*, 2001), and two-tone complexes varying in their relative onset times (Sugimoto *et al.*, 2002).

The spatially distributed activity representing VOT and aspiration noise supports the emerging concept that complex sounds are encoded by temporally precise responses within a network of neuronal populations distributed across tonotopically organized A1 (e.g., Creutzfeldt *et al.*, 1980; Wang *et al.*, 1995; Gehr *et al.*, 2000; Nagarajan *et al.*, 2002; see also Petersen *et al.*, 2002). These distributed responses are time-locked to transient elements embedded within the sounds, and synchronized with each other (Creutzfeldt *et al.*, 1980; Wang *et al.*, 1995). This property may help bind together activity engaged in the processing of speech, which has multiple spectral components activating widespread areas of A1. A temporally precise population coding method is relatively resistant to the addition of background noise (Wang *et al.*, 1995), a prerequisite for speech perception in noisy environments.

Integration of the distributed activity in A1 likely occurs in secondary auditory cortex. Primate A1 projects to multiple surrounding areas, including those located on the lateral surface of the superior temporal gyrus (see Morel *et al.*, 1993; Kaas *et al.*, 1999). MUA and PSTHs recorded from lower lamina 3 are especially relevant to this process, as pyramidal neurons from this depth project to these secondary fields (Galaburda and Pandya, 1983; Jones *et al.*, 1995). Lateral belt areas preferentially respond to vocalizations and noise bursts (Rauschecker *et al.*, 1995; Tian *et al.*, 2001). This characteristic implies that their activation requires integrated activity emanating from multiple frequency-specific areas of A1. Thus, neuronal activity in these areas could unite the different attributes of speech sounds represented in A1. The synchronized responses within A1 would be well-suited for driving these integration cells of nonprimary cortex (Eggermont, 1994, 2000a; deCharms and Merzenich, 1996), leading to a more holistic phonetic representation that incorporates many of the acoustic variables modifying the differential perception of voiced and unvoiced stop consonants.

The ability to ascribe MUA and CSD sinks and sources

to specific neuronal populations and events is central to our interpretations. Response latencies obtained in the present study and attributed to responses of TC fibers and A1 neurons are similar to those previously reported (e.g., Allon *et al.*, 1981; Phillips and Hall, 1990; Heil, 1997). We found peak response latencies of 12–13 ms following stimulus onset in A1 populations within lower lamina 3, and responses in TC fibers located 0.8–1.2 mm below the cortical recordings that peaked 2–3 ms earlier. A conduction velocity of about 0.4 m/s is calculated if responses were only derived from distal segments of TC afferents. This is much slower than calculated velocities of distal TC afferent inputs into A1 (Metherate and Cruikshank, 1999), supporting the conclusion that lower lamina 3 MUA predominantly reflects activity in postsynaptic cortical cells, while deeper MUA is mainly a reflection of distal, TC fiber input. This conclusion is further supported by cross correlations that reveal an average lag time of 1–5 ms between A1 activity and that in the medial geniculate (Creutzfeldt *et al.*, 1980; Miller *et al.*, 2001; see also Usrey, 2002).

The early cortical sink in lower lamina 3 and the later, more superficial sink that we have analyzed are characteristic findings in A1, and likely reflect monosynaptic EPSPs from TC fiber input induced by glutamate and mediated by AMPA/KA receptors, and later polysynaptic EPSPs indicative of intracortical processing, respectively (Metherate and Ashe, 1994; Klink *et al.*, 1999; Metherate and Cruikshank, 1999; Rose and Metherate, 2001; Cruikshank *et al.*, 2002). TC fibers emanating from the main lemniscal (ventral) division of the medial geniculate nucleus ascend through lower cortical laminae and primarily terminate within lamina 4 and lower lamina 3 of A1 (Jones and Burton, 1976; Hashikawa *et al.*, 1995). It is the synaptic activity induced by this input that largely generates the coextensive, early cortical sink (Steinschneider *et al.*, 1992). In contrast, the later, more superficial sink is likely the net result of depolarizations produced by multiple inputs, including those from lamina 4 and contralateral A1 (Mitani *et al.*, 1985; Wallace *et al.*, 1991; Pandya and Rosene, 1993). Furthermore, synaptic events in this laminar region are modulated by cholinergic activity (Metherate and Ashe, 1995; Bandrowski *et al.*, 2001). Thus, the persistence of differential responses reflecting syllable VOT and aspiration noise in the superficial sink offers an opportunity for speech-related activity to interact with learning-related processes. The early current source associated with MUA suppression is also a feature of A1 physiology (Fig. 7), and has been attributed to $GABA_A$ receptor-mediated inhibitory post-synaptic potentials from inhibitory interneurons (Metherate and Ashe, 1994; Metherate and Cruikshank, 1999; Cruikshank *et al.*, 2002) and Ca^{++}-gated K^+ channel-mediated afterhyperpolarization (Eggermont, 2000b). This functional refractory period appears to amplify differential processing of VOT by suppressing responses to voicing for short VOT stimuli, suggesting that multiple factors including peripheral and subcortical processing, A1 inhibitory circuitry, and intrinsic membrane properties are responsible for shaping the temporal response patterns.

Ultimately, the utility of studying speech processing in animals must be assessed by how well the responses model those observed in humans. Multiple studies in humans using noninvasive recording techniques have observed temporal response patterns evoked by syllables and their nonspeech analogs similar to those in monkey A1 (e.g., Kaukoranta *et al.*, 1987; Mäkelä *et al.*, 1988; Kuriki *et al.*, 1995; Sharma and Dorman, 1999). One study of note has reported that the physiological boundary for the N1 component of the scalp-recorded AEP did not reliably predict the voiced/voiceless distinction between stop consonants varying in both VOT and place of articulation (Sharma *et al.*, 2000). While there was concordance between the perceptual and physiological boundaries for the /ba/–/pa/ contrast, the perceptual boundary for /ga/–/ka/ increased to greater than 40 ms while the physiological boundary remained between 20 and 40 ms. This report, however, needs to be interpreted with caution. N1 is a later, composite wave that reflects activation of multiple auditory cortical fields, each with its own capacity to follow temporal features of complex sounds (Steinschneider *et al.*, 1999; Fishman *et al.*, 2001). It is therefore difficult to extrapolate the relevance of short latency, A1 temporal response patterns for phonetic processing from N1 properties. Finally, the two studies examining speech-evoked activity recorded directly from Heschl's gyrus identified response patterns similar to those seen in the monkey, with short VOT syllables eliciting a ''single on'' response and those with a prolonged VOT eliciting ''double on'' responses (Liégeois-Chauvel *et al.*, 1999; Steinschneider *et al.*, 1999). The latter study, using the same synthetic syllables as those in the present study, showed differential responses with a boundary of between 20- and 40-ms VOT that paralleled subject perception. Clearly, additional research must address this problem in greater detail. At the very least, the primate data serve as a valuable model for human physiological activity, whose specific relevance to phonetic perception remains to be determined by future investigations.

V. SUMMARY AND CONCLUSIONS

Representation of stop consonant–vowel syllables with variable VOT is examined in A1 of awake monkeys to test the hypothesis that temporal response patterns differentially reflecting onsets of consonant release and voicing are maintained in large-scale population activity. In low BF areas, syllables with VOTs of 0 and 20 ms evoke a peak of activity time-locked to consonant release, while those with VOTs of 40 and 60 ms elicit an additional peak time-locked to voicing onset. Aspiration noise is represented by sustained increases in neural activity within higher BF areas. While similar patterns are seen in responses emanating from thalamocortical afferents, there is an accentuation of differential activity induced by intracortical mechanisms. Neural suppression following responses to consonant release may serve as a functional refractory period that diminishes responses to voicing onset for syllables with a short VOT. Physiological responses are compatible with the perceptual boundary of +20 to +40 ms that occurs in many languages. These findings support the hypothesis that the 20- to 40-ms interval represents a natural psychoacoustical boundary utilized for VOT perception, based in turn on determining whether the onsets of consonant release and voicing onset occur sequentially in time. Syn-

chronized responses offer an economical way in which stop consonants can be rapidly represented by auditory cortex using the well-described propensity of A1 for responding to low-frequency acoustic transients. These signature patterns are poised to provide potent inputs to higher order auditory fields involved in further phonetic encoding. Secondary fields likely integrate the activity patterns occurring in disparate tonotopic regions of A1, thereby uniting responses representing the various acoustic cues used for the differential perception of voiced from unvoiced stop consonants.

ACKNOWLEDGMENTS

This research was supported by Grant Nos. DC00657, HD01799, and NS07098. We thank Dr. Charles Schroeder, Dr. David Reser, and Dr. Haftan Eckholdt, and Ms. Shirley Seto and Ms. May Huang for providing excellent intellectual, technical, statistical, and histological assistance.

Ahissar, E., Abeles, M., Ahissar, M., Haidarliu, S., and Vaadia, E. (**1998**). ''Hebbian-like functional plasticity in the auditory cortex of the behaving monkey,'' Neuropharmacology **37**, 633–655.

Aitkin, L. M., and Dunlap, C. W. (**1968**). ''Interplay of excitation and inhibition in the cat medial geniculate body,'' J. Neurophysiol. **31**, 44–61.

Allen, J. S., and Miller, J. L. (**1999**). ''Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllable words,'' J. Acoust. Soc. Am. **106**, 2031–2039.

Allon, N., Yeshurun, Y., and Wollberg, Z. (**1981**). ''Responses of single cells in the medial geniculate body of awake squirrel monkeys,'' Exp. Brain Res. **41**, 222–232.

Bandrowski, A. E., Moore, S. L., and Ashe, J. H. (**2001**). ''Cholinergic synaptic potentials in the supragranular layers of auditory cortex,'' Synapse **41**, 118–130.

Barna, J., Arezzo, J. C., and Vaughan, Jr., H. G. (**1981**). ''A new multicontact array for the simultaneous recording of field potentials and unit activity,'' Electroencephalogr. Clin. Neurophysiol. **52**, 494–496.

Blake, D. T., Strata, F., Churchland, A. K., and Merzenich, M. M. (**2002**). ''Neural correlates of instrumental learning in primary auditory cortex,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 10114–10119.

Boucher, V. J. (**2002**). ''Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates,'' Percept. Psychophys. **64**, 121–130.

Brosch, M., Bauer, R., and Eckhorn, R. (**1997**). ''Stimulus-dependent modulations of correlated high-frequency oscillations in cat visual cortex,'' Cereb. Cortex **7**, 70–76.

Carney, A. E., Widin, G. P., and Viemeister, N. F. (**1977**). ''Noncategorical perception of stop consonants differing in VOT,'' J. Acoust. Soc. Am. **62**, 961–970.

Chen, G.-D., Nuding, S. C., Narayan, S. S., and Sinex, D. G. (**1996**). ''Responses of single neurons in the chinchilla inferior colliculus to consonant-vowel syllables differing in voice onset time,'' Aud. Neurosci. **3**, 179–198.

Chen, G.-D., and Sinex, D. G. (**1999**). ''Effects of interaural time differences on the responses of chinchilla inferior colliculus neurons to consonant-vowel syllables,'' Hear. Res. **138**, 29–44.

Creutzfeldt, O., Hellweg, F.-C., and Schreiner, C. (**1980**). ''Thalamocortical transformation of responses to complex auditory stimuli,'' Exp. Brain Res. **39**, 87–104.

Cruikshank, S. J., Rose, H. J., and Metherate, R. (**2002**). ''Auditory thalamocortical synaptic transmission *in vitro*,'' J. Neurophysiol. **87**, 361–384.

deCharms, R., and Merzenich, M. M. (**1996**). ''Primary cortical representation of sounds by the coordination of action-potential timing,'' Nature (London) **381**, 610–613.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (**1997**). ''Perception of synthetic /ba/–/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*),'' J. Acoust. Soc. Am. **102**, 1891–1897.

Dooling, R. J., Okanoya, K., and Brown, S. D. (**1989**). ''Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced–voiceless distinction,'' Percept. Psychophys. **46**, 65–71.

Eggermont, J. J. (**1994**). ''Neural interaction in cat primary auditory cortex. II. Effects of sound stimulation,'' J. Neurophysiol. **71**, 246–270.

Eggermont, J. J. (**1995a**). ''Representation of a voice onset time continuum in primary auditory cortex of the cat,'' J. Acoust. Soc. Am. **98**, 911–920.

Eggermont, J. J. (**1995b**). ''Neural correlates of gap detection and auditory fusion in cat auditory cortex,'' NeuroReport **6**, 1645–1648.

Eggermont, J. J. (**2000a**). ''Sound-induced synchronization of neural activity between and within three auditory cortical areas,'' J. Neurophysiol. **83**, 2708–2722.

Eggermont, J. J. (**2000b**). ''Neural responses in primary auditory cortex mimic psychophysical, across-frequency-channel, gap-detection thresholds,'' J. Neurophysiol. **84**, 1453–1463.

Eilers, R. E., Gavin, W., and Wilson, W. R. (**1979**). ''Linguistic experience and phonetic perception in infancy: A crosslinguistic study,'' Child Dev. **50**, 14–18.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (**1971**). ''Speech perception in infants,'' Science **171**, 303–306.

Faulkner, A., and Rosen, S. (**1999**). ''Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception,'' J. Acoust. Soc. Am. **106**, 2063–2073.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (**2000a**). ''Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness,'' J. Acoust. Soc. Am. **108**, 235–246.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (**2000b**). ''Complex tone processing in primary auditory cortex of the awake monkey. II. Pitch versus critical band representation,'' J. Acoust. Soc. Am. **108**, 247–262.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., Howard, M. A., and Steinschneider, M. (**2001**). ''Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans,'' J. Neurophysiol. **86**, 2761–2788.

Freeman, J. A., and Nicholson, C. (**1975**). ''Experimental optimization of current source density techniques for anuran cerebellum,'' J. Neurophysiol. **38**, 369–382.

Gaese, B. H., and Ostwald, J. (**2001**). ''Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex,'' J. Neurophysiol. **86**, 1062–1066.

Galaburda, A. M., and Pandya, D. N. (**1983**). ''The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkey,'' J. Comp. Neurol. **221**, 169–184.

Gehr, D. D., Komiya, H., and Eggermont, J. J. (**2000**). ''Neuronal responses in cat primary auditory cortex to natural and altered species-specific call,'' Hear. Res. **150**, 27–42.

Hashikawa, T., Molinari, M., Rausell, E., and Jones, E. G. (**1995**). ''Patchy and laminar termination of medial geniculate axons in monkey auditory cortex,'' J. Comp. Neurol. **362**, 195–208.

Heil, P. (**1997**). ''Auditory cortical onset responses revisited. I. First-spike timing,'' J. Neurophysiol. **77**, 2616–2641.

Hirsh, I. J. (**1959**). ''Auditory perception of temporal order,'' J. Acoust. Soc. Am. **31**, 759–767.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). ''Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?'' J. Acoust. Soc. Am. **109**, 764–774.

Jones, E. G., and Burton, H. (**1976**). ''Areal differences in the laminar distribution of thalamic afferents in cortical fields of the insular, parietal and temporal regions of primates,'' J. Comp. Neurol. **168**, 197–248.

Jones, E. G., Dell'Anna, M. E., Molinari, M., Rausell, E., and Hashikawa, T. (**1995**). ''Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity,'' J. Comp. Neurol. **362**, 153–170.

Jusczyk, P. W., Rosner, B. S., Reed, M. A., and Kennedy, L. J. (**1989**). ''Could temporal order differences underlie 2-month-olds' discrimination of English voicing contrasts?'' J. Acoust. Soc. Am. **85**, 1741–1749.

Kaas, J. H., Hackett, T. A., and Tramo, M. J. (**1999**). ''Auditory processing in primate cerebral cortex,'' Curr. Opin. Neurobiol. **9**, 164–170.

Kaukoranta, E., Hari, R., and Lounasmaa, O. V. (**1987**). ''Responses of the human auditory cortex to vowel onset after fricative consonants,'' Exp. Brain Res. **69**, 19–23.

Kewley-Port, D., Watson, C. S., and Foyle, D. C. (**1988**). ''Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli,'' J. Acoust. Soc. Am. **83**, 1133–1145.

Klinke, R., Kral, A., Heid, S., Tillein, J., and Hartmann, R. (**1999**). ''Recruitment of the auditory cortex in congenitally deaf cats by long-term cochlear electrostimulation,'' Science **285**, 1729–1733.

Kluender, K. R. (**1991**). ''Effects of first formant onset properties on voicing judgments result from processes not specific to humans,'' J. Acoust. Soc. Am. **90**, 83–96.
Kluender, K. R., and Lotto, A. J. (**1994**). ''Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans,'' J. Acoust. Soc. Am. **95**, 1044–1052.
Kluender, K. R., Lotto, A. J., and Jenison, R. L. (**1995**). ''Perception of voicing for syllable-initial stops at different intensities: Does synchrony capture signal voiceless stop consonants?'' J. Acoust. Soc. Am. **97**, 2552–2567.
Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (**1998**). ''Role of experience for language-specific functional mappings of vowel sounds,'' J. Acoust. Soc. Am. **104**, 3568–3582.
Kuhl, P. K. (**1981**). ''Discrimination of speech by nonhuman animals: Basic auditory sensitivities conductive to the perception of speech-sound categories,'' J. Acoust. Soc. Am. **70**, 340–349.
Kuhl, P. K., and Miller, J. D. (**1975**). ''Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants,'' Science **190**, 69–72.
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.
Kuhl, P. K., and Padden, D. M. (**1982**). ''Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques,'' Percept. Psychophys. **32**, 542–550.
Kuriki, S., Okita, Y., and Hirata, Y. (**1995**). ''Source analysis of magnetic field responses from the human auditory cortex elicited by short speech sounds,'' Exp. Brain Res. **104**, 144–152.
Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (**1975**). ''VOT discrimination by four- to six- and-a-half-month-old infants from Spanish environments,'' J. Exp. Child Psychol. **20**, 215–225.
Le Prell, C. G., Niemiec, A. J., and Moody, D. B. (**2001**). ''Macaque thresholds for detecting increases in intensity: Effects of formant structure,'' Hear. Res. **162**, 29–42.
Liégeois-Chauvel, C., de Graaf, J. B., Laguitton, V., and Chauvel, P. (**1999**). ''Specialization of left auditory cortex for speech perception in man depends on temporal coding,'' Cereb. Cortex **9**, 484–496.
Lisker, L. (**1975**). ''Is it VOT or a first-formant transition detector?'' J. Acoust. Soc. Am. **57**, 1547–1551.
Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.
Lotto, A. J., and Kluender, K. R. (**2002**). ''Synchrony capture hypothesis fails to account for effects of amplitude on voicing perception,'' J. Acoust. Soc. Am. **111**, 1056–1062.
Lu, T., Liang, L., and Wang, X. (**2001**). ''Temporal and rate representations of time-varying signals in the auditory cortex of awake monkeys,'' Nat. Neurosci. **4**, 1131–1138.
Madler, C., Keller, I., Schwender, D., and Pöppel, E. (**1991**). ''Sensory information processing during general anaesthesia: Effect of isoflurane on auditory evoked neuronal oscillations,'' Br. J. Anaesth. **66**, 81–87.
Mäkelä, J. P., Hari, R., and Leinonen, L. (**1988**). ''Magnetic responses of the human auditory cortex to noise/square wave transitions,'' Electroencephalogr. Clin. Neurophysiol. **69**, 423–430.
Mäkelä, J. P., Karmos, G., Molnár, M., Csépe, V., and Winkler, I. (**1990**). ''Steady-state responses from the cat auditory cortex,'' Hear. Res. **45**, 41–50.
McGee, T., Kraus, N., King, C., and Nicol, T. (**1996**). ''Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe,'' J. Acoust. Soc. Am. **99**, 3606–3614.
Merzenich, M. M., and Brugge, J. F. (**1973**). ''Representation of the cochlear partition on the superior temporal plane of the macaque monkey,'' Brain Res. **50**, 275–296.
Merzenich, M. M., Tallal, P., Peterson, B., Miller, S., and Jenkins, W. M. (**1999**). ''Some neurological principles relevant to the origins of—and the cortical plasticity-based remediation of—developmental language impairments,'' in *Neuronal Plasticity: Building A Bridge From The Laboratory To The Clinic*, edited by J. Grafman and Y. Christen (Springer, Berlin), pp. 169–187.
Metherate, R., and Ashe, J. H. (**1994**). ''Facilitation of an NMDA receptor-mediated EPSP by paired-pulse stimulation in rat neocortex via depression of GABAergic IPSPs,'' J. Physiol. (London) **481**, 331–348.
Metherate, R., and Ashe, J. H. (**1995**). ''Synaptic interactions involving acetylcholine, glutamate, and GABA in rat auditory cortex,'' Exp. Brain Res. **107**, 59–72.
Metherate, R., and Cruikshank, S. J. (**1999**). ''Thalamocortical inputs trigger a propagating envelope of gamma-band activity in auditory cortex *in vitro*,'' Exp. Brain Res. **126**, 160–174.
Miller, L. M., Escabi, M. A., Read, H. L., and Schreiner, C. E. (**2001**). ''Functional convergence of response properties in the auditory thalamocortical system,'' Neuron **32**, 151–160.
Mitani, A., Shimokouchi, M., Itoh, K., Nomura, S., Kudo, M., and Mizuno, N. (**1985**). ''Morphology and laminar organization of electrophysiologically identified neurons in the primary auditory cortex in the cat,'' J. Comp. Neurol. **235**, 430–447.
Morel, A., Garraghty, P. E., and Kaas, J. H. (**1993**). ''Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys,'' J. Comp. Neurol. **335**, 437–459.
Müller-Preuss, P., and Mitzdorf, U. (**1984**). ''Functional anatomy of the inferior colliculus and the auditory cortex: Current source density analyses of click-evoked potentials,'' Hear. Res. **16**, 133–142.
Nagarajan, S. S., Cheung, S. W., Bedenbaugh, P., Beitel, R. E., Schreiner, C. E., and Merzenich, M. M. (**2002**). ''Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex,'' J. Neurophysiol. **87**, 1723–1737.
Nelken, I., Prut, Y., Vaadia, E., and Abeles, M. (**1994**). ''Population responses to multifrequency sounds in the cat auditory cortex: One- and two-parameter families of sounds,'' Hear. Res. **72**, 206–222.
Ohlemiller, K. K., Jones, L. B., Heidbreder, A. F., Clark, W. W., and Miller, J. D. (**1999**). ''Voicing judgments by chinchillas trained with a reward paradigm,'' Behav. Brain Res. **100**, 185–195.
Pandya, D. N., and Rosene, D. L. (**1993**). ''Laminar termination patterns of thalamic, callosal, and association afferents in the primary auditory area of the rhesus monkey,'' Exp. Neurol. **119**, 220–234.
Parker, E. M. (**1988**). ''Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgments of tone-onset simultaneity,'' J. Acoust. Soc. Am. **83**, 1597–1607.
Petersen, R. S., Panzeri, S., and Diamond, M. E. (**2002**). ''Population coding in somatosensory cortex,'' Curr. Opin. Neurobiol. **12**, 441–447.
Phillips, D. P., and Hall, S. E. (**1990**). ''Response timing constraints on the cortical representation of sound time structure,'' J. Acoust. Soc. Am. **88**, 1403–1411.
Pisoni, D. B. (**1977**). ''Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops,'' J. Acoust. Soc. Am. **61**, 1352–1361.
Pisoni, D. B., and Lazarus, J. H. (**1974**). ''Categorical and noncategorical modes of speech perception along the voicing continuum,'' J. Acoust. Soc. Am. **55**, 328–333.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). ''Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants,'' J. Exp. Psychol. Hum. Percept. Perform. **8**, 297–314.
Rauschecker, J. P., Tian, B., and Hauser, M. (**1995**). ''Processing of complex sounds in the macaque nonprimary auditory cortex,'' Science **268**, 111–114.
Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (**1993**). ''Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys,'' J. Neurosci. **13**, 87–103.
Repp, B. H. (**1979**). ''Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants,'' Lang Speech **22**, 173–189.
Rose, H. J., and Metherate, R. (**2001**). ''Thalamic stimulation largely elicits orthodromic, rather than antidromic, cortical activation in an auditory thalamocortical slice,'' Neuroscience **106**, 331–340.
Rotman, Y., Bar-Yosef, O., and Nelken, I. (**2001**). ''Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex,'' Hear. Res. **152**, 110–127.
Schreiner, C. E. (**1998**). ''Spatial distribution of responses to simple and complex sounds in the primary auditory cortex,'' Audiol. Neuro-Otol. **3**, 104–122.
Schroeder, C. E., Tenke, C. E., Givre, S. J., Arezzo, J. C., and Vaughan, Jr., H. G. (**1990**). ''Laminar analysis of bicuculline-induced epileptiform activity in area 17 of the awake macaque,'' Brain Res. **515**, 326–330.
Sharma, A., and Dorman, M. F. (**1999**). ''Cortical auditory evoked potential correlates of categorical perception of voice-onset time,'' J. Acoust. Soc. Am. **106**, 1078–1083.
Sharma, A., Marsh, C. M., and Dorman, M. F. (**2000**). ''Relationship between N1 evoked potential morphology and the perception of voicing,'' J. Acoust. Soc. Am. **108**, 3030–3035.
Sinex, D. G., and McDonald, L. P. (**1988**). ''Average discharge rate repre-

sentation of voice onset time in the chinchilla auditory nerve,'' J. Acoust. Soc. Am. **83**, 1817–1827.
Sinnott, J. M., and Adams, F. S. (**1987**). ''Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts,'' J. Acoust. Soc. Am. **82**, 1539–1547.
Sinnott, J. M., and Brown, C. H. (**1997**). ''Perception of the American English liquid/ra–la/contrast by humans and monkeys,'' J. Acoust. Soc. Am. **102**, 588–602.
Soli, S. D. (**1983**). ''The role of spectral cues in the discrimination of voice onset time differences,'' J. Acoust. Soc. Am. **73**, 2150–2165.
Steinschneider, M., Tenke, C., Schroeder, C., Javitt, D., Simpson, G. V., Arezzo, J. C., and Vaughan, Jr., H. G. (**1992**). ''Cellular generators of the cortical auditory evoked potential initial component,'' Electroencephalogr. Clin. Neurophysiol. **84**, 196–200.
Steinschneider, M., Schroeder, C., Arezzo, J. C., and Vaughan, Jr., H. G. (**1994**). ''Speech-evoked activity in primary cortex: effects of voice onset time,'' Electroencephalogr. Clin. Neurophysiol. **92**, 30–43.
Steinschneider, M., Schroeder, C. E., Arezzo, J. C., and Vaughan, Jr., H. G. (**1995**). ''Physiologic correlates of the voice onset time (VOT) boundary in primary auditory cortex (A1) of the awake monkey: Temporal response patterns,'' Brain Lang **48**, 326–340.
Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., and Arezzo, J. C. (**1998**). ''Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception,'' J. Acoust. Soc. Am. **104**, 2935–2955.
Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A., III (**1999**). ''Temporal encoding of the voice onset time (VOT) phonetic parameter by field potentials recorded directly from human auditory cortex,'' J. Neurophysiol. **82**, 2346–2357.
Steinschneider, M., Volkov, I., Zotova, E., Bakken, H., Fishman, Y., Damasio, H., Arezzo, J. C., and Howard, III, M. (**2000**). ''Differential representation of the voice onset time phonetic parameter within human auditory cortex and a monkey model,'' Soc. Neurosci. Abst. **26**(2), 1969.
Stevens, K. N. (**1981**). ''Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics, and psychoacoustics,'' in *The Cognitive Representation of Speech*, edited by T. Myers, J. Laver, and J. Anderson (North-Holland, Amsterdam), pp. 61–74.
Sugimoto, S., Hosokawa, Y., Horikawa, J., Nasu, M., and Taniguchi, I. (**2002**). ''Spatial focusing of neuronal responses induced by asynchronous two-tone stimuli in the guinea pig auditory cortex,'' Cereb. Cortex **12**, 506–514.
Summerfield, Q. (**1981**). ''Articulatory rate and perceptual constancy in phonetic perception,'' J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.
Summerfield, Q., and Haggard, M. (**1977**). ''On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants,'' J. Acoust. Soc. Am. **62**, 435–448.
Talwar, S. K., and Gerstein, G. L. (**2001**). ''Reorganization in awake rat auditory cortex by local microstimulation and its effect on frequency-discrimination behavior,'' J. Neurophysiol. **86**, 1555–1572.
Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (**2001**). ''Functional specialization in rhesus monkey auditory cortex,'' Science **292**, 290–293.
Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (**1997**). ''Central auditory system plasticity: Generalization to novel stimuli following listening training,'' J. Acoust. Soc. Am. **102**, 3762–3773.
Usrey, W. M. (**2002**). ''The role of spike timing for thalamocortical processing,'' Curr. Opin. Neurobiol. **12**, 411–417.
Utman, J. A. (**1998**). ''Effects of local speaking rate context on the perception of voice-onset time in initial stop consonants,'' J. Acoust. Soc. Am. **103**, 1640–1653.
Vaughan H. G., Jr. and Arezzo, J. C. (**1988**). ''The neural basis of event-related potentials,'' in *Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 3, Human Event-Related Potentials*, edited by T. W. Picton (Elsevier, Amsterdam), pp. 45–96.
Wallace, M. N., Kitzes, L. M., and Jones, E. G. (**1991**). ''Intrinsic inter- and intralaminar connections and their relationship to the tonotopic map in cat primary auditory cortex,'' Exp. Brain Res. **86**, 527–544.
Wang, X., Merzenich, M. M., Beitel, R., and Schreiner, C. E. (**1995**). ''Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics,'' J. Neurophysiol. **74**, 2685–2706.
Weinberger, N. M. (**1997**). ''Learning-induced receptive field plasticity in the primary auditory cortex,'' Semin. Neurosci. **9**, 59–67.
Zurita, P., Villa, A. E. P., de Ribaupierre, Y., de Ribaupierre, F., and Rouiller, E. M. (**1994**). ''Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions,'' Neurosci. Res. (NY) **19**, 303–316.

The effects of a high-frequency suppressor on tuning curves and derived basilar-membrane response functions

Ifat Yasin[a] and Christopher J. Plack
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England

(Received 9 August 2002; revised 11 March 2003; accepted 14 April 2003)

Forward-masked psychophysical tuning curves were obtained using a fixed, low-level signal at a frequency of 4 kHz, and masker frequencies of 2.0, 2.5, 3.0, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75, 5.0, and 5.5 kHz, at masker–signal gaps of 20, 30, 40, 60, 80, and 100 ms. An adaptive two-interval, two alternative forced-choice (2I-2AFC) procedure was used to obtain the masker level at threshold. This procedure was repeated with the addition of a 4.75-kHz suppressor at 50 or 60 dB SPL, gated with the masker. Tuning curves were broader, and estimates of compression and gain from derived input/output functions were decreased in the presence of a suppressor as compared to the no-suppressor condition. The results are consistent with physiological results, which show that suppression leads to a broadening of tuning curves and a partial linearization of the midlevel portion of the basilar-membrane input/output function. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1579003]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [NFV]

I. INTRODUCTION

Both psychological (Nelson and Schroder, 1997; Oxenham and Plack, 1997; Plack and Oxenham, 1998; Nelson *et al.*, 2001) and physiological (Rhode, 1971; Yates, 1990; Ruggero *et al.*, 1992, 1997) studies reveal the response of the basilar membrane (BM) in the cochlea to be highly compressive for a signal at characteristic frequency (CF; the frequency at which the BM response to a low-level sound is greatest). The physiological BM input/output (I/O) function (a plot of the BM vibration as a function of the input sound level) shows that for low-level sounds below about 40 dB SPL, the amplification is linear; the gradient of the I/O function (the response growth) at these sound levels is approximately 1 dB/dB. For midlevel sounds (40–80 dB SPL), the increase in response growth is less than the increase in the input sound level; the response is nonlinear and grows at a rate of approximately 0.2 dB/dB (Sellick *et al.*, 1982; Robles *et al.*, 1986; Yates *et al.*, 1990; Ruggero *et al.*, 1992, 1997). The BM response to midlevel sounds is described as compressive. For frequency components well below CF (for example, an octave below CF) the BM response grows linearly, at least at high CFs (Robles *et al.*, 1986; Ruggero *et al.*, 1992). Another manifestation of this nonlinear processing is suppression (Sellick *et al.*, 1982; Ruggero *et al.*, 1992; Cooper, 1996; Geisler and Nuttall, 1997). Suppression may be attributed to a biasing of the OHC response to the saturating portion of the hair cell's I/O function (a plot of the cell's receptor potential as a function of the acoustic pressure input), thereby reducing the extent of the active mechanical feedback to the BM (e.g., Geisler *et al.*, 1990). Both compression and suppression appear to play a significant role in the processing of complex stimuli such as speech (Stelmachowicz *et al.*, 1982; Rhode and Recio, 2001).

Psychophysical estimates of compression may be derived from either simultaneous- or forward-masking studies. In simultaneous masking the masker and signal are presented together. In forward masking the masker is presented before the signal. In a simultaneous-masking situation the masker may act to suppress (reduce) the BM response to the signal (Ruggero *et al.*, 1992; Nelson and Schroder, 1997; Delgutte, 1990a). Consequently, simultaneous-masking studies of compression (Stelmachowicz *et al.*, 1987) have consistently underestimated the magnitude of compression when compared to forward-masking (Oxenham and Plack, 1997; Nelson *et al.*, 2001) and physiological (Sellick *et al.*, 1982; Ruggero *et al.*, 1992) studies. The contribution of suppression is avoided in forward-masking studies by temporally separating the signal and masker, and in physiological studies by measuring the compressive response of the BM for different levels of a signal presented alone.

Some psychophysical studies of compression have used forward masking to measure the growth of maskability (the masker level required to just mask the signal as a function of signal level) (Oxenham and Plack, 1997; Nelson *et al.*, 2001). The main drawback of this design is that, as the signal level is increased, there is a spread of signal excitation. The signal may then be detected from a region of the BM which has a CF higher than that of the signal (off-frequency listening) (Leshowitz and Wightman, 1971; Johnson-Davies and Patterson, 1979; O'Loughlin and Moore, 1981a). Recently, Nelson *et al.* (2001) have described a method of estimating compression in which the signal is fixed at a low level, thereby minimizing the effects of off-frequency listening. In this design, masker levels required to achieve a constant amount of forward masking are measured as a function of masker–signal delay, plotted as a temporal masking curve (TMC). As the masker–signal gap is increased, a higher masker level is required to mask the signal. As the on-frequency masker level increases it will move from the linear to the compressive region of the BM I/O function. For an on-frequency masker, the increase in masker level represents

[a]Electronic mail: iyasin@essex.ac.uk

the increase required to maintain a constant amount of forward masking as the gap is increased, as well as to compensate for any compression. For an off-frequency masker well below the signal frequency, the masker increase represents only the increase in masker level required to maintain a constant amount of forward masking at the signal place. A comparison of the gradients of the TMCs for the off- and on-frequency masker conditions provides an estimate of the compression of the on-frequency masker.

For a given masker–signal gap, it can be assumed that the internal response of the masker (at the signal place) required to mask the signal, for either an off- or an on-frequency masker is the same, though the external masker levels will differ. Assuming again that the BM response to a masker well below the signal frequency is linear, the lower-frequency masker level is an estimate of the BM response level at the signal place (give or take an additive constant). For a given masker–signal gap, therefore, the lower-frequency masker level (output level) can be plotted against the on-frequency masker level (input level) to produce an estimate of the I/O function for the on-frequency masker (Nelson *et al.*, 2001). This psychophysical I/O function is comparable to the physiological BM I/O function (Cooper and Yates, 1994).

Houtgast (1974) demonstrated how the forward-masked threshold of a pure-tone signal might be reduced (unmasked) by the addition of a second pure-tone signal (the suppressor) at another frequency, presented simultaneously with the pure-tone masker (the suppressee). The suppressor reduces the response of the masker, at the signal location on the BM, making the signal easier to detect. In addition, physiological studies suggest that suppression may reduce compression by linearizing the BM I/O function (Robles *et al.*, 1987; Ruggero *et al.*, 1992). Ruggero *et al.* used signals at CFs of 8–10 kHz at levels between 30 and 80 dB SPL. The BM I/O functions for these signals were compressive; a given increase in signal level produced a relatively smaller increase in BM vibration. Suppressor tones were presented at frequencies higher or lower than CF. In the presence of a suppressor, a signal at CF produced a smaller response on the BM. However, the suppression was restricted to low to medium signal levels, resulting in a more linear BM I/O function; a given increase in signal level now produced a near-equivalent increase in BM vibration.

The aim of the present study was twofold. First, to use the TMC technique to map the range of frequencies adjacent to CF that are affected by compression. Second, to estimate the effects of suppression on the BM I/O function. O'Loughlin and Moore (1981b) compared forward-masked psychophysical tuning curves (PTCs) with and without a suppressor, and found that masker levels for masker frequencies within about 250 Hz of the signal frequency were increased, resulting in a broadening of the PTC tip. The present study measured this effect over a wider range of masker frequencies. Duifhuis (1980) demonstrated a greater magnitude of suppression for a suppressor higher, rather than lower, in frequency than the signal. The present study measures the effect of suppression on a range of masker levels for two levels of a high-frequency suppressor.

II. METHOD

A. Conditions

The levels of a sinusoidal forward masker required to mask a 4-kHz signal were obtained for masker frequencies of 2.0, 2.5, 3.0, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75, 5.0, and 5.5 kHz, and masker–signal gaps (t_{ms}) of 20, 30, 40, 60, 80, and 100 ms. Thresholds were obtained in the presence of a notched noise at a passband spectrum level of −10 dB, to restrict further the available off-frequency excitation (Oxenham and Plack, 1997).[1] The signal was presented at 35 dB SPL, which was 10 dB above signal threshold in the notched noise for two of the subjects (AM and IY) and 5 dB above for one of the subjects (PP). The maskers were presented in each of three suppressor conditions; no suppressor, with a 50-dB SPL suppressor, and with a 60-dB SPL suppressor. The suppressor levels were chosen such that they were at least about 10 dB below the level of a 4.75-kHz masker required to mask the signal. The sinusoidal suppressor was presented at 4.75 kHz, since preliminary studies had shown that a suppressor at this frequency produced the greatest suppression of a 4-kHz masker. For two of the subjects (AM and PP), the three suppressor conditions were presented as a fully randomized set for each t_{ms}. For one of the subjects (IY), the no-suppressor and 50-dB suppressor conditions were run first, with the 60-dB suppressor condition presented later.

B. Stimuli

The 4-kHz signal had no steady-state portion, and 4-ms raised-cosine onset and offset ramps. Both the masker and suppressor had a steady-state duration of 40 ms and 2-ms raised-cosine onset and offset ramps. The masker and suppressor were gated together. t_{ms} was specified as the silent time interval between the zero-voltage points of the envelopes at the end of the masker and at the beginning of the signal. The notched noise had a center frequency of 4 kHz, a notch width of 800 Hz (3-dB downpoints), and attenuation slopes of 90 dB/octave. The noise was gated on 50 ms before the onset of the masker and gated off 50 ms after the offset of the signal. All stimuli were digitally generated by a PC using a sampling rate of 48 kHz. Stimuli were output via a soundcard with 24-bit resolution. Antialiasing was provided by built-in filters. The stimuli were presented to the right channel of a set of Sennheiser HD 580 headphones. The headphone input came directly from the output of the soundcard DAC.

C. Procedure

Subjects were tested individually while seated in an IAC double-walled sound-attenuating booth. The stimuli were presented to the subjects' right ears. A 2I-2AFC adaptive procedure was used to determine the masker level at which the subject would achieve 70.7% correct on the psychometric function (Levitt, 1971). A block of trials began with the presentation of a light on a computer-simulated response box. Subjects started a block of trials by pressing a start key. The length of each observation interval was indicated by a light on the response box in one of two rectangles. The interstimu-

lus interval was 500 ms. On each trial, the masker (and suppressor, if present) was presented in both intervals. The signal was presented at random in one of the intervals. The task within each trial was to select the signal interval. Subjects responded by pressing the appropriate response key. After a response, visual feedback was provided by the presentation of a colored light. Initially, the masker was decreased after an incorrect response and increased after every two consecutive correct responses. A reversal was counted every time the masker level changed direction. The masker level was varied in steps of 4 dB for the first four reversals. For the following 12 reversals the step size was reduced to 2 dB. The levels for the last 12 reversals were averaged to obtain the threshold value. In this way, an estimate of threshold was obtained from each block of trials. If the standard deviation of the estimated threshold value was greater than 7 dB, the block of trials was repeated. Once performance was stable (as judged by the experimenter), the mean threshold was calculated from the final four estimates of threshold, with the most deviant threshold discarded. If, however, the mean of the last four thresholds had a standard deviation greater than 2.5 dB, then the mean of the final six thresholds was taken, with the most deviant threshold discarded. For the masker condition of 2 kHz, the mean of the final eight thresholds was taken with the most deviant threshold discarded. The maximum masker level that could be produced by the system was 102 dB SPL. The system peak clipped at this level, producing distortion at the headphone output. If a masker level of 102 dB SPL was reached within a block of trials, the estimate of threshold from that block was discarded.

D. Subjects

Three subjects were tested. One was the author IY and the other two subjects were paid for their services. Absolute thresholds were obtained for signal frequencies of 0.25, 0.5, 1, 2, 4, and 8 kHz, using the adaptive procedure described earlier, but in this case the signal level was varied in a 2-down, 1-up procedure. The signal had a steady-state duration of 200 ms and 10-ms raised-cosine onset and offset ramps. Absolute thresholds at 4 kHz for subjects AM, PP, and IY were −4.2, −5.0, and 8.2 dB SPL, respectively. Absolute thresholds for the 4-kHz signal used in the experiment were 5.1, 4.8, and 13.5 dB SPL, respectively. Thresholds for the 4-kHz signal used in the experiment in the presence of notched noise alone were 26.2, 31.6, and 23.4 dB SPL, respectively.

III. RESULTS

A. PTCs and estimated filter shapes

Although there were individual differences in the overall masker levels at threshold, the form of the PTCs was consistent across listeners. Mean PTCs for the no-suppressor and 50- and 60-dB suppressor conditions are shown in the left column of Fig. 1. For clarity the error bars have been omitted (the average standard errors for PP, IY, and AM were 2.9, 2, and 2.8 dB, respectively). Each panel presents a set of PTCs. Each PTC represents the level of a masker of a given frequency required to mask a 4-kHz signal for a given t_{ms}.

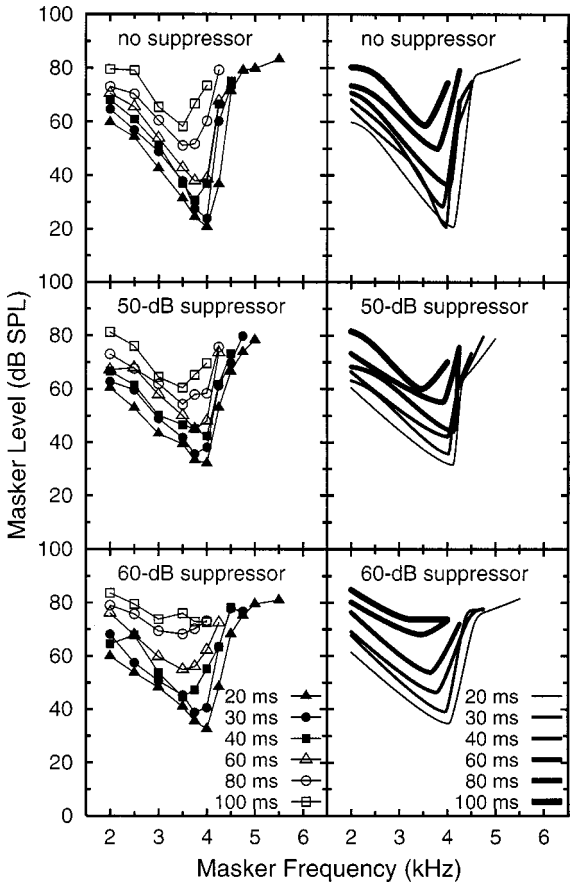


FIG. 1. The left column presents the mean PTCs with and without a 4.75-kHz suppressor at either 50 or 60 dB. Each PTC represents the masker level at threshold for a 4-kHz signal as a function of masker frequency. PTCs for t_{ms} of 20, 30, 40, 60, 80, and 100 ms are shown. Filled triangles, circles, and squares represent PTCs for 20, 30, and 40 ms, respectively. Open triangles, circles, and squares represent PTCs for 60, 80, and 100 ms, respectively. Error bars are omitted for clarity. The right column presents the roex (p,w,t) filter shape fits to mean PTCs without, and with a 50- or 60-dB SPL suppressor. Each panel presents PTCs for 20, 30, 40, 60, 80, and 100 ms represented by lines of increasing thickness.

Each data point represents the mean threshold across three subjects. In some cases, particularly for longer t_{ms}, complete PTCs could not be obtained as the masker level reached the output limit of the equipment.

The mean PTC data were fitted with a rounded exponential (roex) function with three free parameters (p,w,t) (Patterson *et al.*, 1982). The roex (p,w,t) filter shape comprises two rounded exponentials; the first characterizes the passband of the filter, the second, the tail outside the passband

$$W(g)=(1-w)(1+pg)\exp(-pg) + w(1+tg)\exp(-tg), \quad (1)$$

where $W(g)$ is the noise power at the output of the auditory filter, p determines the slope of the tip of the filter, t determines the slope of the tail of the filter (the rate of fall of the second exponential term), and w determines when the second exponential takes over from the first. The choice of the roex (p,w,t) filter function was determined by the observation that many of the mean PTCs have low- and high-frequency

TABLE I. Roex (p,w,t) filter shape predicted values for L_{BF}, BF, pu, pl, wu, wl, tu, tl, and Q_{10} and bandwidth estimates for PTCs, at t_{ms} of 20 to 100 ms, with no-suppressor, and with a 50- or 60-dB suppressor. The root-mean-squared deviations for the filter shape fit to the experimental data are also shown.

		Roex filter shape parameters										
Suppressor condition	PTC per t_{ms} (ms)	pl	wl	tl	pu	wu	tu	L_{BF} (dB)	BF (kHz)	rms	10-dB Bandwidth (kHz)	Q_{10}
No suppressor	20	26.16	1.01E−04	0.01	169.74	2.76E−06	2.76E−06	20.70	4.13	1.04	0.71	5.83
	30	54.01	1.73E−02	16.82	255.18	1.07E−03	55.33	20.61	3.97	0.09	0.35	11.25
	40	34.34	1.30E−02	14.09	141.09	4.00E−04	30.45	28.39	3.91	0.03	0.55	7.08
	60	23.58	2.75E−04	0.00	195.88	5.39E−04	131.73	36.25	4.05	0.11	0.75	5.43
	80	20.41	3.91E−03	0.00	74.62	2.17E−03	74.61	49.96	3.79	0.98	0.91	4.15
	100	24.02	6.45E−03	0.00	41.53	5.84E−03	367.50	58.63	3.53	1.58	0.89	3.99
50-dB suppressor	20	17.37	1.80E−04	39.90	237.26	1.01E−03	27.90	31.60	4.13	0.93	0.99	4.16
	30	20.14	1.41E−03	0.02	207.28	7.98E−03	41.33	35.76	4.03	1.25	0.85	4.72
	40	15.40	1.29E−02	408.76	132.33	3.27E−02	46.00	42.17	4.01	1.16	1.13	3.56
	60	19.24	3.38E−03	0.00	335.68	9.70E−04	463.21	44.23	4.14	1.94	0.89	4.63
	80	12.49	1.51E−04	25.28	77.73	1.70E−04	7.58	55.08	3.91	1.43	1.40	2.80
	100	20.20	5.38E−03	0.00	25.08	4.25E−03	129.99	59.99	3.46	0.83	1.18	2.92
60-dB suppressor	20	16.65	7.99E−05	23.57	91.10	1.27E−04	9.02	34.65	4.04	1.33	0.93	4.32
	30	17.82	3.54E−09	10.47	108.59	1.47E−04	0.00	39.07	3.96	1.04	0.93	4.28
	40	15.47	1.92E−04	23.13	51.54	7.22E−04	0.00	46.40	3.77	1.05	1.21	3.11
	60	15.91	5.92E−04	15.91	36.57	2.50E−04	36.58	53.94	3.64	0.46	1.25	2.92
	80	10.35	7.52E−05	8.54	15.02	1.54E−04	11.66	68.06	3.44	0.64	2.13	1.62
	100	10.45	2.49E−04	21.42	0.00	3.05E−04	43.79	73.77	3.22	1.18	⋯	⋯

tails. In such a case, the use of two rounded exponentials on each skirt of the filter generates a more accurate fit to the data. Since the filter is assumed to be asymmetric, p, w, and t were allowed to have different values for the low- (pl,wl,tl) and high- (pu,wu,tu) frequency slopes of the filter. These different functions are used for frequencies below and above best frequency (BF), respectively. (The BF of the PTC is defined as the frequency of the masker most effective at masking the signal.) Here, g is the deviation of the evaluation frequency (f) from the BF, normalized by dividing by the BF of the filter

$$g=|f-\mathrm{BF}|/\mathrm{BF}. \tag{2}$$

A least-squares minimization procedure was implemented to fit the roex (p,w,t) filter shape to the PTC data. The roex filter shape was converted to dB attenuation and inverted. To find the best fit, the filter was shifted in frequency (i.e., BF was varied) and in level (i.e., L_{BF}, the level of the fitted PTC at BF, was varied). Values of parameters pl, wl, tl, pu, wu, tu, BF, and L_{BF} were found that minimized the root-mean-squared (rms) deviation between the experimental and predicted data values.

The predicted values for L_{BF}, BF, pl, pu, wl, wu, tl, tu, as well as the rms deviations for the filter shape fit to the experimental data, are presented in Table I. The roex filter fits to the experimental data are shown in the right column of Fig. 1. The roex filter functions generally provide a good fit to the data, reflecting the main features of the mean PTC data in the presence of a suppressor as compared to the PTC data in the absence of a suppressor.

The 10-dB bandwidths of the filters were found by subtracting the frequency value of the low- and high-frequency filter slopes at which the difference in attenuation relative to the filter tip was 10 dB. The sharpness of tuning was found by dividing this bandwidth estimate by BF, to derive a value of Q_{10}. BFs and Q_{10} values relative to the tip level of the PTC for the no-suppressor and 50- or 60-dB suppressor filter fits are shown in Fig. 2. For all three suppressor conditions, the BF generally shifts to a lower frequency as t_{ms} increases (increases in t_{ms} correspond to a higher masker level, L_{BF}, at the tip of the PTC). At high masker levels the peak of the masker excitation pattern shifts towards the basal region of the BM (basalward shift) (e.g., McFadden, 1986). One consequence of this basalward shift is that the 4-kHz masker is now less effective than a lower frequency masker at masking the 4-kHz signal.

The Q_{10} estimates of the 50- and 60-dB suppressor PTCs are generally less than those for the no-suppressor PTCs, consistent with a broadening of the auditory filter (a reduction in frequency selectivity) in the presence of a suppressor. Q_{10} estimates decrease gradually with increases in the

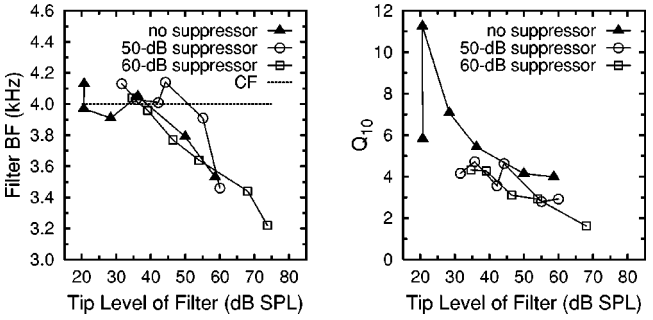


FIG. 2. The left panel presents roex filter fit values for BF as a function of tip level for t_{ms} of 20–100 ms. The dashed line represents CF for reference. The right panel presents roex filter fit values for Q_{10} as a function of tip level for t_{ms} of 20–100 ms. Each panel presents data without a suppressor (filled triangles), with a 50-dB suppressor (open circles) or a 60-dB suppressor (open squares).

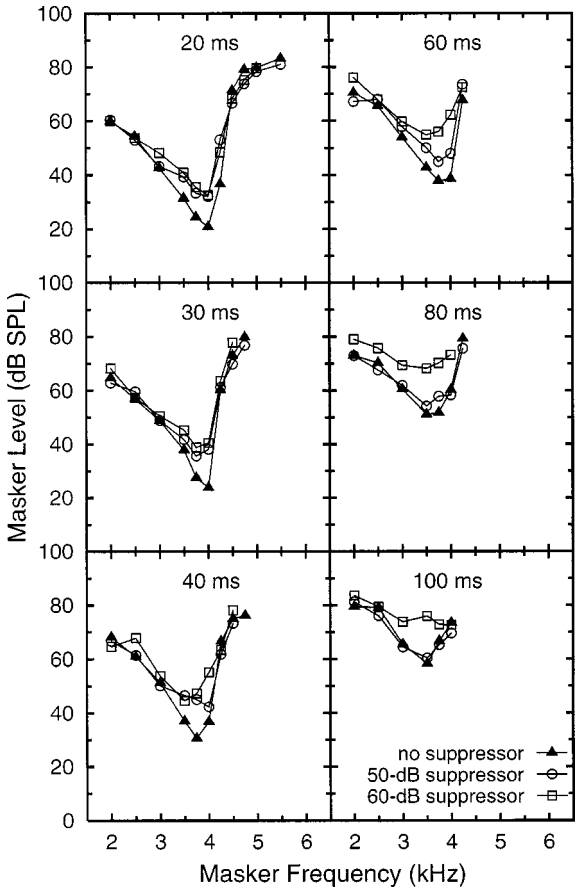


FIG. 3. Each panel presents a set of mean PTCs for a given t_{ms} of 20, 30, 40, 60, 80, or 100 ms. Each set of PTCs consists of a PTC without a suppressor (filled triangless), with a 50-dB suppressor (open circles) or a 60-dB suppressor (open squares).

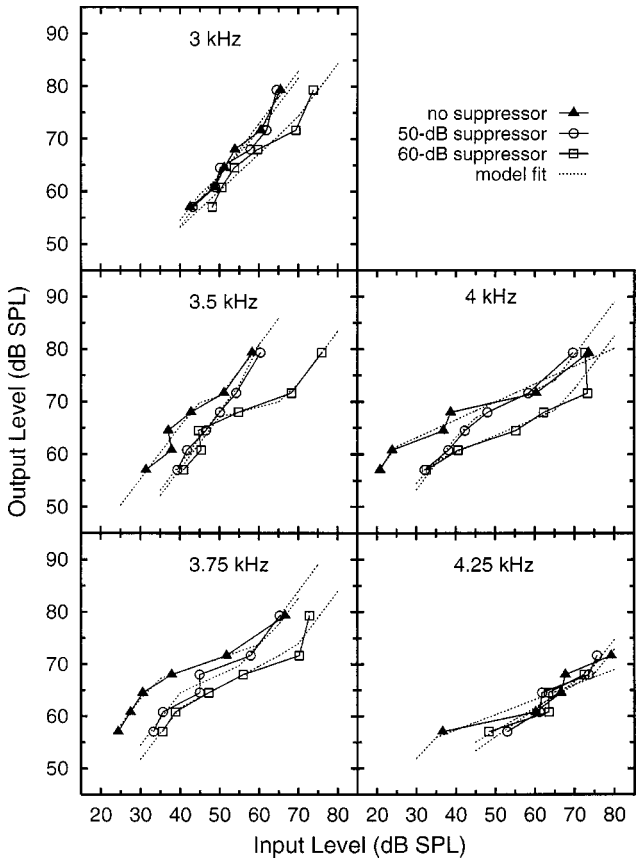


FIG. 4. I/O functions derived from mean PTC data. Each panel shows the output masker levels (an average of 2- and 2.5-kHz masker levels) as a function of the input masker levels for masker frequencies of 3, 3.5, 3.75, 4, and 4.25 kHz. Each panel presents three I/O functions; with no suppressor (filled triangles), with a 50-dB suppressor (open circles), and with a 60-dB suppressor (open squares). Dotted lines are model fits (functions 1, 2, and 3) to the experimental data. Error bars are omitted for clarity.

masker level (apart from a sharp increase for the no-suppressor PTC as the t_{ms} is increased from 20–30 ms). The dynamic range for the 60-dB suppressor PTC at a t_{ms} of 100 ms was too small to allow an estimate of Q_{10}.

B. Range of frequencies affected by suppression

To establish the range of masker frequencies affected by the suppressor, the no-suppressor and 50- or 60-dB suppressor PTCs are depicted together in Fig. 3. Each panel presents a no-suppressor PTC (filled triangles), a 50-dB suppressor PTC (unfilled circles), and a 60-dB suppressor PTC (unfilled squares) for a given t_{ms}.

The 50- and 60-dB suppressor PTCs are elevated, with respect to the no-suppressor PTC, primarily for masker frequencies close to the signal frequency (3.5–4.25 kHz). In contrast, there is minimal difference between the 2-kHz masker level as a function of t_{ms}, with, or without, a suppressor. Compared to the no-suppressor PTC, the 60-dB suppressor PTC is elevated for masker frequencies in the range 3.5–4.25 kHz for all t_{ms}, whereas, for the same frequency range, the 50-dB suppressor PTC is only elevated for t_{ms} up to 60 ms. At larger t_{ms} (80 and 100 ms) the 50-dB SPL suppressor and no-suppressor PTCs overlap. For example, for the 4-kHz masker, for t_{ms} of 20 and 30 ms, the amount of suppression afforded by a 50- or 60-dB SPL suppressor is approximately the same. As the t_{ms} increases there is a greater difference between the suppressors. At t_{ms} of 80 and 100 ms, the 50-dB suppressor no longer contributes to suppression of the masker, whereas the 60-dB suppressor still produces a large elevation of the masker level for masker frequencies in the range 3.5–4.25 kHz.

C. Derived I/O functions

The magnitude of compression of a given masker frequency at the signal location can be estimated by the slope of the plot of the mean masker level for the off-frequency maskers (2+2.5 kHz), against the masker level of the given masker frequency (Fig. 4). The 2- and 2.5-kHz masker levels were averaged to reduce the margin of error in estimation of the linear response at CF. The response to both these input frequencies should be linear at CF (Ruggero *et al.*, 1992). Each point on the I/O function represents a pairing of the mean 2+2.5-kHz masker level, with the on- or off-frequency masker level, for a given t_{ms}. Each panel presents three I/O functions for a given masker frequency; an I/O function for the no-suppressor, and the 50- or 60-dB suppressor conditions. For the no-suppressor condition the I/O functions for maskers close to the signal frequency are shallow, indicating compression. For these frequencies the addition of a suppressor shifted the I/O function to the right.

TABLE II. The model predictions of G, c, BP_1, and BP_2, for masker frequencies of 3, 3.5, 3.75, 4, and 4.25 kHz, with no-suppressor, and with a 50- or 60-dB suppressor. Values for BP_1 and BP_2, which exceeded the minimum or maximum input masker levels, are omitted. Values for G for which BP_1 was less than the minimum input masker level are omitted.

		Model parameters			
Input frequency (kHz)	Suppressor condition	G (dB)	c (dB/dB)	BP_1 (dB)	BP_2 (dB)
3	No-suppressor	14.44	0.29	46.15	48.38
	50-dB suppressor	13.33	0.28	53.56	56.05
	60-dB suppressor		0.70		69.40
3.5	No-suppressor	25.32	0.24	44.34	50.36
	50-dB suppressor		0.36		57.18
	60-dB suppressor	17.16	0.19	49.91	66.91
3.75	No-suppressor	33.29	0.26	33.59	61.59
	50-dB suppressor	24.45	0.34	40.19	56.12
	60-dB suppressor	21.74	0.40	39.75	69.48
4	No-suppressor	36.63	0.34	24.77	
	50-dB suppressor	23.28	0.35	42.93	63.88
	60-dB suppressor	24.48	0.32	33.97	66.31
4.25	No-suppressor		0.29		
	50-dB suppressor		0.53		73.96
	60-dB suppressor		0.44		

For the purposes of estimating compression ratios, a model was used which applied a three-function fit to the data: a linear low-level region (function 1), a compressive midlevel region (function 2), and a linear high-level region (function 3). The three functions were joined by two breakpoints, a lower breakpoint (BP_1) joining functions 1 and 2, and an upper breakpoint (BP_2) joining functions 2 and 3

$$L_{out} = L_{in} + G \quad (L_{in} \leqslant BP_1), \tag{3}$$

$$L_{out} = cL_{in} + k_1 + G \quad (BP_1 < L_{in} \leqslant BP_2), \tag{4}$$

$$L_{out} = L_{in} + k_2 + G \quad (L_{in} > BP_2), \tag{5}$$

where G is the gain (dB), c is the slope of the compressive region (dB/dB), $k_1 = BP_1(1-c)$, $k_2 = BP_2(c-1) + k_1$, L_{in} (physical level, dB SPL) and L_{out} (level of BM response, dB).

A least-squares minimization procedure was implemented using MATLAB's ''fminsearch'' routine; the values of the free parameters (G, c, BP_1, and BP_2) were found that minimized the rms deviations between the experimental values from the derived I/O functions and the predicted values.

The predicted values for G, c, BP_1, and BP_2 are shown in Table II. The dotted lines in Fig. 4. represent the model predictions for the I/O functions. Generally, the model accounts well for the experimental data. The slopes from function 2 of the fitted I/O functions, for maskers between 3 and 4.25 kHz, are shown in Table II. It is important to note that in cases in which BP_1 and BP_2 are similar the I/O function can be assumed to be linear across most of the input range. Although there was variability in individual conditions, there was some evidence that suppression reduced the amount of compression. The average compression exponents for the no-suppressor and 50- and 60-dB suppressor conditions for masker frequencies close to the signal frequency (3.75–4.25 kHz) were 0.30, 0.41, and 0.39, respectively.

The predicted values of G are shown in Table II. If BP_1 is greater than the minimum input masker level, then G is an estimate of the difference between the maximum gain applied to the masker frequency of interest and the average gain applied to the 2- and 2.5-kHz maskers at the 4-kHz place. High values of G are associated with high gain for the cochlear amplifier. A decrease in G is characterized by a shift of the I/O function to higher input masker levels. In all cases in which G was a meaningful estimate of gain, the addition of a suppressor decreased G.

The possibility that confusion and cuing may have contributed to the shape of the with-suppressor PTCs in this study is tested in the Appendix; cuing of the masker offset by the suppressor appears to play a limited role in elevation of masker levels.

IV. DISCUSSION

A. PTCs and estimates of compression without a suppressor

Overall, the shape of the forward-masked PTCs (Fig. 1) is in agreement with earlier psychophysical data which show that high-frequency slopes tend to be steeper than low-frequency slopes (Vogten, 1974, 1978; Small and Busse, 1980; Kidd and Feth, 1981; O'Loughlin and Moore, 1981b; Moore and Glasberg, 1982; McFadden and Yama, 1983; Nelson and Freyman, 1984; Jesteadt and Norton, 1985; Nelson, 1991). In general, the Q_{10} estimates of the PTCs decrease as the masker–signal gap increases, consistent with a broadening of the auditory filter with an increase in level (Moore and Glasberg, 1987; Nelson and Freyman, 1984). The BF of the PTCs shifts to a lower frequency at high masker levels (corresponding to longer masker–signal gaps). Comparable intensity-dependent shifts of the most effective stimulus frequency to lower frequencies have also been observed for BM iso-velocity tuning curves and BM iso-intensity functions (Robles *et al.*, 1986; Ruggero *et al.*, 1997), neural tuning curves (Sachs and Abbas, 1974), and psychophysical masking patterns (McFadden and Yama, 1983; McFadden, 1986; Moore *et al.*, 2002). The results are consistent with an intensity-dependent migration of the peak of the BM traveling wave towards the base of the cochlea (McFadden, 1986).

The results show a flattening of the high-frequency slope (high-frequency plateau) of the PTC, for the shortest masker–signal gap. High-frequency plateaus have been identified for BM iso-velocity tuning curves (Rhode, 1971; Wilson and Johnstone, 1975; Robles *et al.*, 1985, 1986; Cooper and Rhode, 1992; Nuttall and Dolan, 1996; Ruggero *et al.*, 1997; Narayan *et al.*, 1998), as well as in the cochlear microphonics (Dallos and Cheatham, 1971). Although it has been suggested that plateaus in physiological data may be a result of either acoustic trauma (Gummer and Johnstone, 1984), or fluid-depth artifacts associated with laser velocimetry (Cooper and Rhode, 1996; Ruggero *et al.*, 1997), these explanations cannot account for plateaus evident in the data from undamaged preparations (Robles *et al.*, 1986; Ruggero *et al.*, 1997) or from Mössbauer measurements (Robles *et al.*, 1985, 1986). It is noteworthy that although high-frequency plateaus appear to be a prevalent feature of physiological data, there is little indication of the phenomenon in psychophysical studies, with the exception of some data from Nelson (1991). A possible reason for this discrepancy may be that high stimulus levels are often limited by the output of audiometric equipment in psychophysical data collection, predominantly to protect the listener from uncomfortable sound levels. However, this limitation does not explain why high-frequency plateaus are not often reported in neural data.

Data presented in Table II indicate that compression at the 4-kHz CF in response to input frequencies at 3.75 and 4.25 kHz is approximately 0.3 dB/dB (a compression ratio of 3.3:1). Moore *et al.* (1999) report compression estimates of 0.5 for a 4-kHz signal, although the range of compression estimates for their subjects ranged from 0.29–0.55.

A simple comparison of compressive estimates from different studies is problematic for three main reasons. First, compression estimates have been obtained at different signal frequencies, 6 kHz (Oxenham and Plack, 1997), 0.375, 0.75, 1.5, and 3 kHz (Hicks and Bacon, 1999), 2, 4, and 6 kHz (Moore *et al.*, 1999) and 1 kHz (Nelson *et al.*, 2001). An across-frequency comparison is confounded by the fact that some physiological data suggest BM compression to be less compressive at low as compared to high frequencies (Yates, 1990; Rhode and Cooper, 1996). However, recent psychophysical data suggest equivalent compression at low and high CFs (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003). Second, off-frequency listening may have led to an underestimation of the actual compression at the signal location for some studies (Moore *et al.*, 1999). Third, the range of input levels across which compression is estimated differs across studies. For instance, Oxenham and Plack (1997) calculated compression for a limited range of input levels from 50–80 dB SPL, while Moore *et al.* (1999) estimated compression over a wider range. Similarly, compression estimates from rate-level functions have been calculated for different ranges of discharge rates (Cooper and Yates, 1994). It can be said, however, that the current estimates are broadly consistent with those from previous studies.

The present data suggest that the 3.75- and 4.25-kHz maskers are compressed to a similar extent. The value of *c* for the 4-kHz masker I/O function is slightly greater than for adjacent masker frequencies. However, the compressive portion of the I/O function for the 4-kHz masker extended much further beyond the input range. Since the values for BP_1 and BP_2 (Table II) are similar for masker frequencies of 3 and 3.5 kHz, the I/O functions can be assumed to be almost linear for these frequencies. In direct BM measurements there appears to be a gradual linearization of the response for signal frequencies lower than CF (Ruggero *et al.*, 1997; Russell and Nilsen, 1997; Rhode and Recio, 2001). It was not possible to gain any information regarding compression for maskers higher in frequency than 4.25 kHz because of the output limitations mentioned before. Some physiological data suggest that the BM responses to a signal above CF may be at least as compressive as the response for an on-frequency signal (Sellick *et al.*, 1982; Robles *et al.*, 1986; Cooper and Rhode, 1992; Murugasu and Russell, 1995; Nuttall and Dolan, 1996; Ruggero *et al.*, 1997; Russell and Nilsen, 1997), at least for signal frequencies up to about 1/3 octave above CF (Ruggero *et al.*, 1997; Rhode and Recio, 2001). Compression thereafter appears to decrease rapidly with increasing frequency (Ruggero *et al.*, 1997). It is not obvious from the present data whether the compressive region extends approximately uniformly either side of the signal location (Russell and Nilsen, 1997) or whether the compressive region is basal to the signal location (Neely and Kim, 1986).

There appears to be some return to linearity at high levels for some masker frequencies close to the signal frequency. Some psychophysical data are also suggestive of linearity at high levels for off-frequency growth of maskability (Oxenham and Plack, 1997) and derived I/O functions (Nelson *et al.*, 2001). Some BM I/O functions for a signal at CF approach linearity at high signal levels (Ruggero and Rich, 1991; Ruggero *et al.*, 1992), whereas other data show no return to linearity (Murugasu and Russell, 1995; Ruggero *et al.*, 1997). A return to linearity appears to be dependent on the technique employed to measure the BM response; the effect appears to be more prevalent in interferometry data (Ruggero and Rich, 1991; Ruggero *et al.*, 1992) than in data obtained using the Mössbauer technique (Sellick *et al.*, 1982; Robles *et al.*, 1986).

B. PTCs with a suppressor

In the presence of a suppressor, the masker levels for masker frequencies close to CF are raised, while masker levels at frequencies well removed from CF remain relatively unchanged, in accordance with previous studies (O'Loughlin and Moore, 1981b; Ruggero *et al.*, 1992). The combined effect of this is to broaden the frequency tuning about CF. The Q_{10} values in the presence of a suppressor were generally smaller than those of PTCs obtained without. Forward-masked PTCs in general provide a higher estimate of frequency selectivity than simultaneous-masked PTCs (Moore, 1978). It appears that in the presence of a suppressor, the bandwidth of the forward-masked PTC resembles the wider bandwidths found using simultaneous masking (Wightman *et al.*, 1977; Moore, 1978; Javel *et al.*, 1983; Fahey and Allen, 1985; Ruggero *et al.*, 1992). The present results show that for masker–signal gaps of 20–80 ms, Q_{10} estimates

without a suppressor are on average 40% and 49% larger than those with a 50- or 60-dB suppressor, respectively. Wightman *et al.* (1977) and Moore (1978) reported Q_{10} estimates to be twice as large in forward- as compared to simultaneous-masked PTCs. However, the latter two results may have been influenced by confounding factors such as off-frequency listening (Patterson, 1976; Johnson-Davies and Patterson, 1979). Studies using notched noise to restrict off-frequency listening have found forward-masked PTCs to have 3-dB bandwidths values about 17% less than those estimated using simultaneous masking (Moore and Glasberg, 1981; Moore *et al.*, 1987).

C. Linearization of the I/O function in the presence of a suppressor

Previous studies have shown a shift to higher signal levels (reduced gain) in the presence of a suppressor for neural rate level (Sachs, 1969; Abbas and Sachs, 1976; Delgutte, 1990b) and synchronization I/O functions (Javel *et al.*, 1983). BM I/O functions also show a shift to higher signal levels (Robles *et al.*, 1987; Ruggero *et al.*, 1992; Rhode and Recio, 2001). Furthermore, the amount of suppression increases with increases in suppressor level (Duifhuis, 1980), observable as an increase in the magnitude of the shift of the I/O function (Costalupes *et al.*, 1987; Javel, 1981; Ruggero *et al.*, 1992; Rhode and Recio, 2001). With the addition of a 50-dB suppressor, there is a shift of the I/O function to higher input masker levels (a reduction in gain), except for the 4.5-kHz masker, and linearization of the midportion of the function. An increase in suppressor level to 60 dB does not appear to produce a simple additional shift in the I/O function or a clear linearization of the midportion of the function. (With a 60-dB suppressor the highest compression value corresponds to the 3.5-kHz masker.) One reason for this may be that since the 60-dB suppressor was close to the excitatory region of the tuning curve for all subjects at short masker–signal gaps, it may have contributed to masking of the signal. Masking by the 60-dB suppressor would reduce the masker levels required, and shift the derived I/O function to the left (lower input levels) at short masker–signal gaps.

It has been suggested that suppression operates by affecting the internal response to the stimulus, such that the effect is analogous to that of attenuating the stimulus in level. Sachs and Abbas (1976) originally introduced the idea of attenuation in their phenomenological model of neural rate-intensity functions. According to an attenuation model, a suppressor would produce a constant horizontal shift of the I/O function to higher signal levels. A similar effect is suggested by the data of Javel *et al.* (1978) and Abbas (1978). However, there appear to be some discrepancies between rate- and synchrony-based measures of suppression. The implication is that the shift of neural I/O functions in the presence of a suppressor observed in some studies may not actually reflect operation in accordance with an attenuation-based model of suppression (Delgutte, 1990b).

Present results indicate that in the presence of a suppressor, the I/O function not only shifts to higher masker levels, but also linearizes to some extent. A linearization as well as a shift has been reported for BM (Ruggero *et al.*, 1992; Rhode and Recio, 2001) and neural (Javel *et al.*, 1983) I/O functions. A linearization of the I/O function suggests that suppression may operate by decreasing the nonlinear gain applied to the stimulus (Geisler and Sinex, 1980; Javel, 1981; Geisler, 1985; Deng and Geisler, 1985; Robles *et al.*, 1986; Cheatham and Dallos, 1989; Nuttall and Dolan, 1993; Rhode and Recio, 2001). A suppressor may achieve a reduction in gain by driving the compressive nonlinearity at the hair cell level into saturation (Geisler *et al.*, 1990). The effect of this would be to reduce the BM response at the location corresponding to the frequency of the suppressee. The present results appear to be consistent with an explanation of suppression based on gain reduction.

V. CONCLUSIONS

(i) In the presence of a high-frequency suppressor, the tip of the forward-masked PTC for a 4-kHz signal was raised (higher masker levels for maskers near the signal frequency), resulting in broader tuning. Q_{10} estimates without a suppressor were approximately 40% and 49% larger than Q_{10} estimates with a 50- and 60-dB suppressor respectively.

(ii) The effects of the suppressor were limited to masker frequencies between approximately 3 and 4 kHz.

(iii) In the presence of a suppressor, the derived I/O function was shifted to higher masker levels (less gain), and the compression estimate derived from the gradient of the midlevel portion of the function was reduced slightly.

ACKNOWLEDGMENTS

The authors thank Vit Drga for theoretical and technical support and Dr. Deb Fantini for helpful criticisms and suggestions. We also thank two anonymous reviewers for their helpful comments. The research was supported by EPSRC Grant GR/N07219 and by the ESRC.

APPENDIX: CONTROL FOR CUING

In a typical forward-masking signal-detection task, a subject can use a number of cues to detect the signal. If the masker is of a frequency close to, or the same as, the signal, there may be an absence of a pitch distinction to cue detection of the signal (Terry and Moore, 1977). The situation is further complicated if the masker–signal gap is short, in which case there is also no temporal distinction. If the masker is of a different frequency or the masker–signal gap is long enough, a pitch or temporal distinction is available as a cue to the subject, and the signal is relatively easier to detect.

If the situation is such that masker and signal frequencies are similar and the masker–signal gap is short, then this may lead to confusion as to when the masker ends and the signal starts. Evidence from Moore and Glasberg (1982) suggests that in such a situation, signal detection may be improved by gating a cue (like the suppressor tone used here) with the masker. The suppressor may resolve confusion to some extent by acting as a temporal marker, indicating when

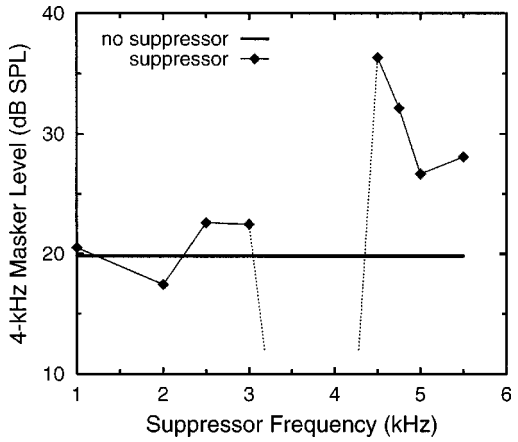


FIG. 5. The level of a 4-kHz masker required to mask a 4-kHz signal as a function of suppressor frequency, for a 50-dB suppressor. The solid line represents the 4-kHz masker level required to mask the 4-kHz signal in the absence of a suppressor. A suppressor at a frequency of 3.5, 3.75, 4, or 4.25 kHz masked the signal on its own, such that the 4-kHz masker level from the adaptive procedure was very low. These values are suggested by the dotted lines in the figure.

the masker ends and the signal starts (Neff, 1985, 1986). The implications of the effects of confusion and cuing as applied to the present study are, first, that confusion may lead to a lowering of masker levels for frequencies close to and at CF, artificially enhancing the tip of the PTC. Second, addition of the suppressor may resolve this confusion, enhancing signal detection, leading to an elevation of the masker levels adjacent to, and at CF, thereby apparently flattening the tip of the PTC. In order to separate the elevation of masker levels attributable to suppression from that attributable to cuing, the present study was undertaken. The study is based on the supposition that a cue, by virtue of its description as a temporal marker, may be expected to exert its effect independently of frequency. Specifically, if the suppressor were acting as a cue then it should be equally effective at raising masker levels for a range of suppressor frequencies. Conversely, previous studies (Houtgast, 1974; Shannon, 1976) suggest that a suppressor is most effective at suppressing the internal representation of a suppressee when the suppressor frequency falls within a small region on either side of the PTC. If the suppressor were acting to suppress the internal representation of the masker, then it would be differentially effective in raising masker levels for a given range of suppressor frequencies.

Detection of a 4-kHz signal was measured in the presence of a 4-kHz masker and a 50-dB SPL suppressor at 1.0, 2.0, 2.5, 3.0, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75, 5.0, or 5.5 kHz. The suppressor was gated with the masker and the masker–signal gap was 20 ms. Mean results from two subjects (IY and PP) indicate that only a small range of suppressor frequencies above and below 4 kHz leads to a significant increase in the masker level, as compared to the no-suppressor condition (Fig. 5). Based on the present results it appears unlikely that resolution of confusion by the addition of the suppressor influenced the elevation of masker levels, even at the shortest gaps tested. Rather, the data support a suppressive reduction of the internal masker representation for masker frequencies close to CF (Houtgast, 1974; O'Loughlin and Moore, 1981a; Moore and Glasberg, 1982).

[1]It appears unlikely that the continuous noise (−10-dB spectrum level) would suppress the off-frequency maskers at the masker levels required for deriving the PTCs. The overall power per ERB of the noise at 3 kHz (for example) would be only about 15 dB.

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.

Abbas, P. J. (**1978**). "Effects of stimulus frequency on two-tone suppression," J. Acoust. Soc. Am. **63**, 1878–1886.

Cheatham, M. A., and Dallos, P. (**1989**). "Two-tone suppression in inner hair cell responses," Hear. Res. **40**, 187–196.

Cooper, N. (**1996**). "Two-tone suppression in cochlear mechanics," J. Acoust. Soc. Am. **99**, 3087–3098.

Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea-pig cochlea: Sharp tuning and nonlinearity in the absence of baseline position shifts," Hear. Res. **63**, 163–190.

Cooper, N. P., and Rhode, W. S. (**1996**). "Two-tone suppression in apical cochlear mechanics," Aud. Neurosci. **3**, 123–134.

Cooper, N. P., and Yates, G. K. (**1994**). "Nonlinear input–output functions derived from the responses of guinea-pig cochlear nerve fiber: Variations with characteristic frequency," Hear. Res. **78**, 221–234.

Costalupes, J. A., Rich, N. C., and Ruggero, M. A. (**1987**). "Effects of excitatory and nonexcitatory suppressor tones on two-tone rate suppression in auditory nerve fibers," Hear. Res. **26**, 155–164.

Dallos, P., and Cheatham, M. A. (**1971**). "Travel time in the cochlea and its determination from cochlear-microphonic data," J. Acoust. Soc. Am. **49**, 1140–1143.

Delgutte, B. (**1990a**). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," J. Acoust. Soc. Am. **87**, 791–809.

Delgutte, B. (**1990b**). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Deng, L., and Geisler, C. D. (**1985**). "Changes in the phase of excitator-tone responses in cat auditory nerve fibers by suppressor tones and fatigue," J. Acoust. Soc. Am. **78**, 1633–1643.

Duifhuis, H. (**1980**). "Level effects in psychophysical two-tone suppression," J. Acoust. Soc. Am. **67**, 914–1643.

Fahey, P. F., and Allen, J. B. (**1985**). "Nonlinear phenomena as observed in the ear canal and at the auditory nerve," J. Acoust. Soc. Am. **77**, 599–612.

Geisler, C. D. (**1985**). "Effects of a compressive nonlinearity in a cochlea model," J. Acoust. Soc. Am. **78**, 257–260.

Geisler, C. D., and Nuttal, A. L. (**1997**). "Two tone suppression of basilar-membrane vibrations in the base of the guinea pig cochlea using 'low-side' suppressors," J. Acoust. Soc. Am. **45**, 986–999.

Geisler, C. D., and Sinex, D. G. (**1980**). "Responses of primary auditory fibers to combined noise and tonal stimuli," Hear. Res. **3**, 317–334.

Geisler, C. D., Yates, G. K., Patuzzi, R. B., and Johnstone, B. M. (**1990**). "Saturation of outer hair cell receptor currents causes two-tone suppression," Hear. Res. **44**, 241–256.

Gummer, A. W., and Johnstone, B. M. (**1984**). "Group delay measurements from spiral ganglion cells in the basal turn of the guinea pig," J. Acoust. Soc. Am. **76**, 1388–1400.

Hicks, M., and Bacon, S. P. (**1999**). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," J. Acoust. Soc. Am. **105**, 326–338.

Houtgast, T. (**1974**). "Lateral suppression in hearing," Ph.D. thesis, Free University of Amsterdam.

Javel, E. (**1981**). "Suppression of auditory nerve responses I. Temporal analysis, intensity effects, and suppression contours," J. Acoust. Soc. Am. **69**, 1735–1745.

Javel, E., Geisler, C. D., and Ravindra, A. (**1978**). "Two-tone suppression in auditory nerve of the cat: Rate-intensity and temporal analysis," J. Acoust. Soc. Am. **63**, 1093–1104.

Javel, E., McGee, J., Walsh, E. J., Farley, G. R., and Gorga, M. P. (**1983**). "Suppression of auditory nerve responses. II. Suppression threshold and growth, iso-suppression contours," J. Acoust. Soc. Am. **74**, 801–813.

Jesteadt, W., and Norton, S. J. (**1985**). "The role of suppression in psychophysical measures of frequency selectivity," J. Acoust. Soc. Am. **78**, 365–373.

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,'' Hear. Res. **5**, 49–67.
Leshowitz, B., and Wightman, F. L. (**1971**). ''On-frequency masking with continuous sinusoids,'' J. Acoust. Soc. Am. **49**, 1180–1190.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). ''Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing,'' J. Acoust. Soc. Am. **113**, 951–960.
McFadden, D. (**1986**). ''The curious half octave shift: Evidence for a basalward migration of the travelling-wave envelope with increasing intensity,'' in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, edited by R. J. Salvi, D. Henderson, R. P. Hamernik, and V. Colleti (Plenum, New York), pp. 295–312.
McFadden, D., and Yama, M. F. (**1983**). ''Upward shifts in the masking pattern with increasing masker intensity,'' J. Acoust. Soc. Am. **74**, 1185–1189.
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., Alcantara, J. I., and Glasberg, B. R. (**2002**). ''Behavioral measurement of level-dependent shift in the vibration pattern on the basilar membrane,'' Hear. Res. **163**, 101–110.
Moore, B. C. J., and Glasberg, B. R. (**1981**). ''Auditory filter shapes derived in simultaneous and forward masking,'' J. Acoust. Soc. Am. **70**, 1003–1014.
Moore, B. C. J., and Glasberg, B. R. (**1982**). ''Interpreting the role of suppression in psychophysical tuning curves,'' J. Acoust. Soc. Am. **72**, 1374–1379.
Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.
Moore, B. C. J., Poon, P. W. F., Bacon, S. P., and Glasberg, B. R. (**1987**). ''The temporal course of masking and the auditory filter shape,'' J. Acoust. Soc. Am. **81**, 1873–1880.
Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). ''Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**, 2761–2777.
Murugasu, E., and Russell, I. J. (**1995**). ''Salicylate ototoxicity: The effects on the basilar membrane displacement, cochlear microphonics, neural responses in the basal turn of the guinea pig cochlea,'' Aud. Neurosci. **1**, 139–150.
Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (**1998**). ''Frequency tuning of the basilar membrane and auditory nerve fibers in the same cochleae,'' Science (Washington, DC, U.S.) **282**, 1882–1884.
Neely, S. T., and Kim, D. O. (**1986**). ''A model for active elements in cochlear biomechanics,'' J. Acoust. Soc. Am. **79**, 1472–1480.
Neff, D. L. (**1985**). ''Stimulus parameters governing confusion effects in forward masking,'' J. Acoust. Soc. Am. **78**, 1966–1976.
Neff, D. L. (**1986**). ''Confusion effects with sinusoidal and narrow-band forward maskers,'' J. Acoust. Soc. Am. **79**, 1519–1528.
Nelson, D. A. (**1991**). ''High-level psychophysical tuning curves: Forward masking in normal hearing and hearing-impaired listeners,'' J. Speech Hear. Res. **34**, 1233–1249.
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. (**1997**). ''Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss,'' J. Acoust. Soc. Am. **101**, 2186–2201.
Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). ''A new procedure for measuring peripheral compression in normal-hearing, and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 2045–2064.
Nuttall, A. L., and Dolan, D. F. (**1993**). ''Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig,'' J. Acoust. Soc. Am. **93**, 390–400.
Nuttall, A. L., and Dolan, D. F. (**1996**). ''Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig,'' J. Acoust. Soc. Am. **99**, 1556–1565.
O'Loughlin, B. J., and Moore, B. C. J. (**1981a**). ''Improving psychacoustical tuning curves,'' Hear. Res. **5**, 343–346.
O'Loughlin, B. J., and Moore, B. C. J. (**1981b**). ''Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking,'' J. Acoust. Soc. Am. **69**, 1119–1125.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Patterson, R. D. (**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 threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Plack, C. J., and Drga, V. (**2003**). ''Psychophysical evidence for auditory compression at low characteristic frequencies,'' J. Acoust. Soc. Am. **113**, 1574–1586.
Rhode, W. S. (**1971**). ''Observations of the vibration of the BM in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*,'' Aud. Neurosci. **3**, 101–121.
Rhode, W. S., and Recio, A. (**2001**). ''Multicomponent stimulus interactions observed in basilar-membrane vibration in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **110**, 3140–3154.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1985**). ''Mossbauer measurements of the mechanical response to single-tone and two-tone stimuli at the base of the chinchilla cochlea,'' in *Peripheral Auditory Mechanism*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, Berlin), pp. 121–128.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output function, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1367.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1987**). ''Two-tone suppression in the basilar membrane of the chinchilla,'' in *Neuroscience. The 2nd World Congress of Neuroscience (IBRO)* (Pergamon Oxford).
Ruggero, M. R., Robles, L., and Rich, N. C. (**1992**). ''Two-tone suppression in the BM of the cochlea: Mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of the outer hair cells upon basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2164.
Russell, I. J., and Nilsen, K. E. (**1997**). ''The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane,'' Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.
Sachs, M. B. (**1969**). ''Stimulus-response relation for auditory-nerve fibers: Two-tone stimuli,'' J. Acoust. Soc. Am. **45**, 1025–1036.
Sachs, M. B., and Abbas, P. J. (**1974**). ''Rate versus level functions for auditory-nerve fibers in cats: Tone-burst stimuli,'' J. Acoust. Soc. Am. **56**, 1835–1847.
Sachs, M. B., and Abbas, P. J. (**1976**). ''Phenomenological model for two-tone suppression,'' J. Acoust. Soc. Am. **60**, 1157–1163.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar-membane motion in the guinea-pig using the Mossbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Shannon, R. V. (**1976**). ''Two-tone unmasking and suppression in a forward masking situation,'' J. Acoust. Soc. Am. **59**, 1460–1470.
Small, A. M., and Busse, C. (**1980**). ''Psychophysical tuning curves in forward masking,'' J. Acoust. Soc. Am. Suppl. 1 **67**, S9.
Stelmachowicz, P. G., Small, A. M., and Abbas, P. J. (**1982**). ''Suppression effects for complex stimuli,'' J. Acoust. Soc. Am. **71**, 410–420.
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.
Terry, M., and Moore, B. C. J. (**1977**). '''Suppression effects' in forward-masking,'' J. Acoust. Soc. Am. **62**, 781–784.
Vogten, L. L. M. (**1974**). ''Pure-tone masking: A new result from a new method,'' in *Facts and Models of Hearing*, edited by E. Zwicker and E. Terhardt (Springer, New York), pp. 142–155.
Vogten, L. L. M. (**1978**). ''Low-level pure-tone masking: A comparison of

'tuning curves' obtained with simultaneous and forward masking,'' J. Acoust. Soc. Am. **63**, 1520–1527.
Wightman, F., McGee, T., and Kramer, M. (**1977**). ''Factors influencing frequency selectivity in normal and hearing-impaired listeners,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
Wilson, J. P., and Johnstone, J. R. (**1975**). ''Basilar-membrane and middle-ear vibration in guinea pig measured by capacitive probe,'' J. Acoust. Soc. Am. **57**, 705–723.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''BM nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**, 203–220.

An account of monaural phase sensitivity

Robert P. Carlyon[a)]
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

Shihab Shamma
Electrical and Computer Engineering, Institute for Systems Research, University of Maryland, College Park, Maryland 20742

(Received 2 January 2002; revised 30 March 2003; accepted 7 April 2003)

Listeners can detect phase differences between the envelopes of sounds occupying remote frequency regions, and between the fine structures of partials that interact within a single auditory filter. They are insensitive to phase differences between partials that differ sufficiently in frequency to preclude within-channel interactions. A new model is proposed that can account for all three of these findings, and which, unlike currently popular approaches, does not discard across-channel timing information. Sensitivity is predicted quantitatively by analyzing the output of a cochlear model using a spectro-temporal decomposition inspired by responses of neurons in the auditory cortex, and by computing a distance metric between the responses to two stimuli to be discriminated. Discriminations successfully modeled include phase differences between pairs of bandpass filtered harmonic complexes, and between pairs of sinusoidally amplitude modulated tones, discrimination between amplitude and frequency modulation, and discrimination of transient signals differing only in their phase spectra (''Huffman sequences''). © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1577557]

PACS numbers: 43.66.Ba, 43.66.Nm, 43.64.Bt [NFV]

I. INTRODUCTION

There is now a considerable amount of evidence concerning the circumstances under which listeners can discriminate between two sounds that differ only in their phase spectra. Differences in the relative phase of two simultaneous frequency components that excite separate frequency channels are not detectable, except when they affect the magnitude of the combination tones generated in the cochlea (Buunen and Bilsen, 1974). However, such phase differences can become highly discriminable once those components interact in the output of one or more auditory filters. This simple rule holds, for example, in the discrimination between amplitude modulation (AM) and quasi-frequency modulation (QFM: Mathes and Miller, 1947; Goldstein, 1966) and between harmonics summed in sine versus alternating phase (Patterson *et al.*, 1988; Shackleton and Carlyon, 1994).

The crucial role of within-channel interactions in the discrimination of phase differences between partials of complex tones has inspired auditory models in which between-channel phase information is discarded. For example, Meddis and Hewitt's (1991a, b) ''autocorrelogram'' model calculates a summary estimate of the temporal structure of each auditory filter output separately, and only then sums these individual ''autocorrelograms'' across channels. Because each single-channel autocorrelogram represents an estimate based on tens of milliseconds of activity, fine between-channel phase information is discarded. Similarly, Patterson's ''auditory image'' model (Patterson *et al.*, 1995) contains a ''strobing'' mechanism, which is explicitly designed to remove timing differences between channels. As a result, both models do a good job of accounting for the effect of resolvability on listeners' sensitivity to phase differences between partials of a complex tone.

A problem for models that explicitly discard across-channel timing information is that, for some stimuli, listeners *can* make across-channel phase comparisons. For example, when two groups of higher-numbered harmonics are filtered into separate frequency regions and presented concurrently, listeners can detect small (1–2 ms) shifts in the relative timing of the two groups, even when the outputs of auditory filters that might respond to both groups are masked by noise (Carlyon, 1994; Carlyon and Shackleton, 1994). Another example of sensitivity to across-channel timing differences is that subjects can detect a 30°–60° phase difference between the envelopes of two AM sinusoids, again when the stimuli are presented against a noise background (Strickland *et al.*, 1989; Yost and Sheft, 1989). Listeners' ability to use this type of information is also reflected in the phenomenon of co-modulation masking release (Hall *et al.*, 1984), the size of which depends on the phase relationship between the envelopes of the separate noise bands that constitute the masker (Haggard *et al.*, 1985; McFadden, 1986; Moore and Schooneveldt, 1990). Because ''autocorrelogram'' and ''auditory image'' models analyze timing information on a within-channel basis, they fail to capture these across-channel differences. We will return to this point in more detail in Sec. IV.

It seems clear, then, that a complete account of monaural phase sensitivity should not only explain why subjects *cannot* detect phase differences between resolved partials of complex tones, but also why they *can* detect across-frequency differences between the envelopes of broader-band stimuli. In addition, such an account should be consis-

a)Electronic mail: bob.carylon@mrc-cbu.cam.ac.uk

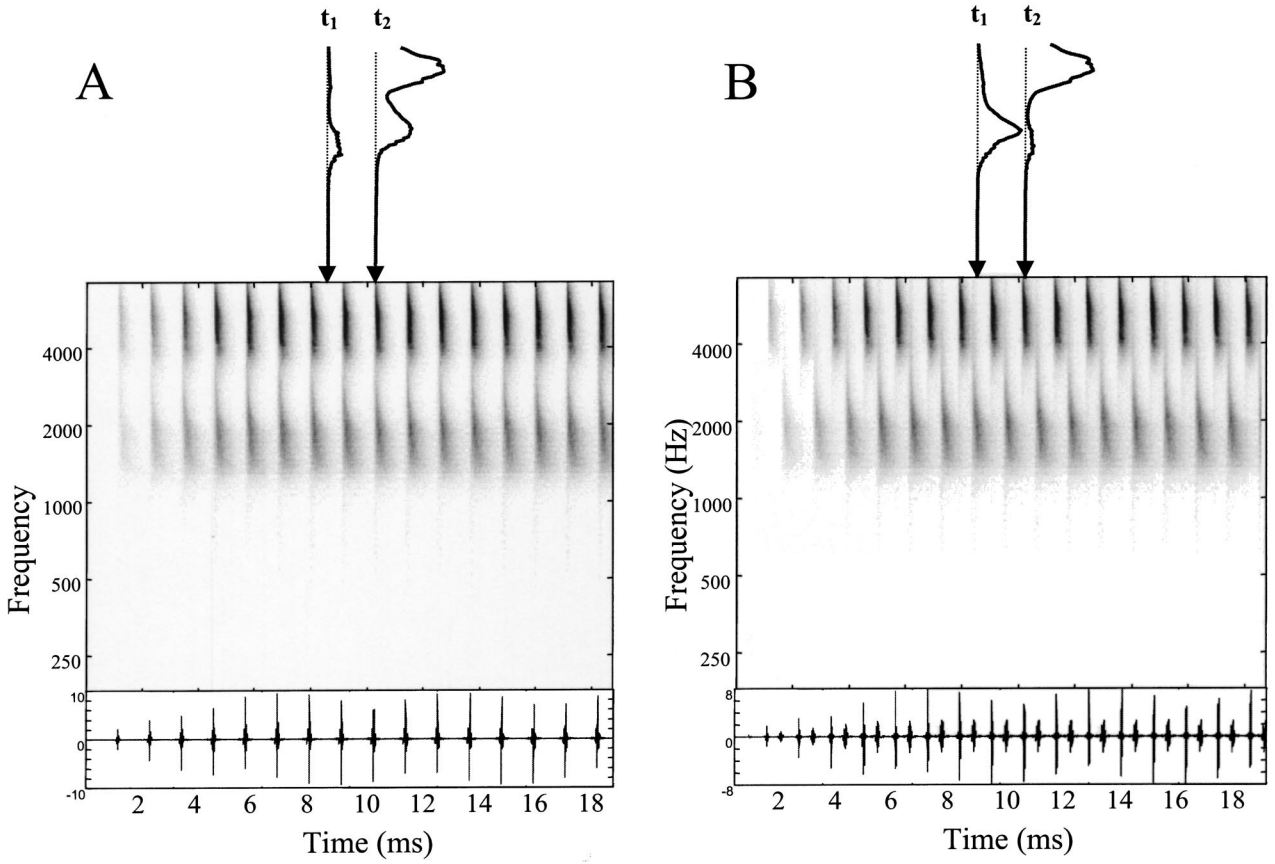


FIG. 1. Auditory spectrograms of unresolved harmonic tones. (a) The auditory spectrogram computed using a model of the early auditory stages. The stimulus is the sum of a pair of unresolved harmonic complexes (see text for details). The stimulus waveform is shown below the spectrogram. Two spectral cross sections at time instants (t_1) and (t_2) are shown above the spectrogram. Note that the two spectral peaks rise and fall roughly in phase with each other. (b) The auditory spectrogram of the 5-ms relatively delayed pair of unresolved harmonic complexes. The stimulus waveform is shown below the spectrogram. Two spectral cross sections at the same time instants (t_1) and (t_2) as in (a) are shown above the spectrogram. Note that the two spectral peaks here rise and fall roughly out-of-phase with each other.

tent with the ability of listeners to detect phase differences between partials that interact within the same auditory filter. The remainder of this article describes a new model of monaural phase sensitivity which achieves these three goals. A crucial difference between the new model and those described previously (Meddis and Hewitt, 1991b; Patterson *et al.*, 1995) is that it does not invoke a loss of across-channel timing information to explain insensitivity to the phase relationship between resolved partials. We start by providing a general overview of the model, which will be illustrated with reference to Carlyon and Shackleton's (1994) experiment on the detection of asynchronies between groups of unresolved harmonics. We shall then proceed to a more detailed description, and, finally, provide quantitative predictions of human phase sensitivity measured using a variety of stimuli and detection tasks.

II. MODEL

A. Intuitive summary

All the experimental results analyzed here can be cast as two-alternative-forced-choice (''2AFC'') discrimination tasks in which two stimuli differ by the value of one parameter. Our goal is to describe a detection model that can predict thresholds for a wide variety of stimuli. The approach we propose is based on a model of sound analysis in the early and cortical stages of auditory processing (Yang *et al.*, 1992; Wang and Shamma, 1994; Lyon and Shamma, 1996; Chi *et al.*, 1999). The model consists of the following computational procedures: (*i*) Compute the AN responses (or the *auditory spectrogram*) to the *standard* and *signal* stimuli. (*ii*) Analyze the spectrograms to represent explicitly their spectral and temporal features. The key requirement here is that the representation directly reflects any changes that subjects can detect. (*iii*) Determine the percentage change in the feature representation at threshold in one task, and use this change to predict threshold in all other tasks.

The first stage of this general approach can be illustrated with reference to Carlyon and Shackleton's (1994) task of detecting an asynchrony between the envelope maxima, or ''pitch pulses,'' in two (concurrently presented) groups of unresolved harmonics. In their experiment, both groups had an F0 of 88 Hz, with one group bandpass filtered between 1375 and 1875 Hz and the other between 3900 and 5400 Hz. The auditory spectrograms (Yang *et al.*, 1992) corresponding to the synchronous and (5-ms) asynchronous delayed complex tones are shown in Figs. 1(a) and (b), respectively. Because the harmonics are unresolved, they interact within the cochlear filters resulting in the 88-Hz modulation evident in the spectrogram responses near (and above) the 1600- and

4600-Hz CF regions—roughly corresponding to the center frequencies of each group of harmonics. Delaying the lower group of harmonics relative to the upper group causes their respective modulations to become progressively less synchronized [Fig. 1(b)], a feature which subjects can detect when the asynchrony exceeds approximately 1.4 ms (Carlyon and Shackleton, 1994).

The stimuli shown in Fig. 1 differ from those used by Carlyon and Shackleton (1994) in that our simulation does not include the background noise that they used. Generally, the use of background noise is essential in experiments on across-channel phase sensitivity (Strickland *et al.*, 1989; Yost and Sheft, 1989; Carlyon, 1994; Carlyon and Shackleton, 1994), in order to mask combination tones (CTs) and within-channel interactions. It is less important for simulations such as ours which do not include CTs and where the absence (or otherwise) of within-channel interactions can be confirmed by visual inspection. In Sec. IV we show that the output of our model in response to the stimuli used by Carlyon and Shackleton is not greatly affected by the addition of background noise.

To quantify the difference in the auditory response to two stimuli, we first characterize the spectral and temporal features of each spectrogram, and then integrate a measure of the difference between these features over the duration of the stimulus. For example, the modulations in Fig. 1(a) are in phase and therefore the spectrum at any time contains two spectral peaks that rise and fall together as shown in the sequence of cross sections at time instants $(\mathbf{t}_1)$ and $(\mathbf{t}_2)$. By comparison, the asynchronized modulations in Fig. 1(b) display very different spectral cross-sections at time instants $(\mathbf{t}_1)$ and $(\mathbf{t}_2)$, exhibiting only one large peak that alternates its position between the two CF regions. Therefore, one way to summarize the difference between the two spectrograms would be to sense the change in their spectral cross sections, and then to integrate this difference over the whole stimulus duration. A simple example of such a measure might be to integrate the absolute difference between the two spectral peaks, which is likely to yield much smaller values for Fig. 1(a) compared to Fig. 1(b).

In the next section we describe a more general formulation to characterize the spectro-temporal features of the auditory spectrograms (and hence their changes with various stimulus manipulations), one which is inspired by human psychoacoustics and animal physiology of the auditory cortex. Detection thresholds of the model will be derived from the results of the above-described task, and then applied to predict performance in a variety of other psychoacoustical tests.

B. Model methodology

1. The auditory spectrogram

Computing the auditory spectrograms (such as those shown in Fig. 1) consists of a sequence of two linear operations.

(i) A frequency analysis stage that consists of a bank of bandpass filters equally spaced on a logarithmic frequency axis. The model employs 24 filters/octave over a 5-octave range. Details of the filter shapes and bandwidth have been described by Yang *et al.* (1992). An example filter, tuned to 1500 Hz, is shown by the dotted lines in Fig. 2(a). It can be seen that, compared to the widely used ''gammatone'' filter (solid line: Patterson *et al.*, 1988), it is less symmetric and is more broadly tuned, with a $Q_{3\,dB}$ of 4 instead of 9.3. Another way of characterizing the sharpness of tuning is in terms of the equivalent rectangular bandwidth (ERB) of the filters; this was 255 Hz for the filter shown in Fig. 2(a) (dotted lines), compared to 188 Hz for the gammatone. Either way, the bandwidth is a constant proportion of center frequency (CF), and the filters are related to each other by a simple dilation of their impulse responses.

(ii) A hair cell stage in which the filter outputs are half-wave rectified, and then low-pass filtered with a time constant of 1 ms, simulating the loss of phase-locking on the auditory nerve (Lyon and Shamma, 1996). Note, however, that this stage contains no compression, and so the calculation of the auditory spectrogram is linear with respect to the total power in the signal.

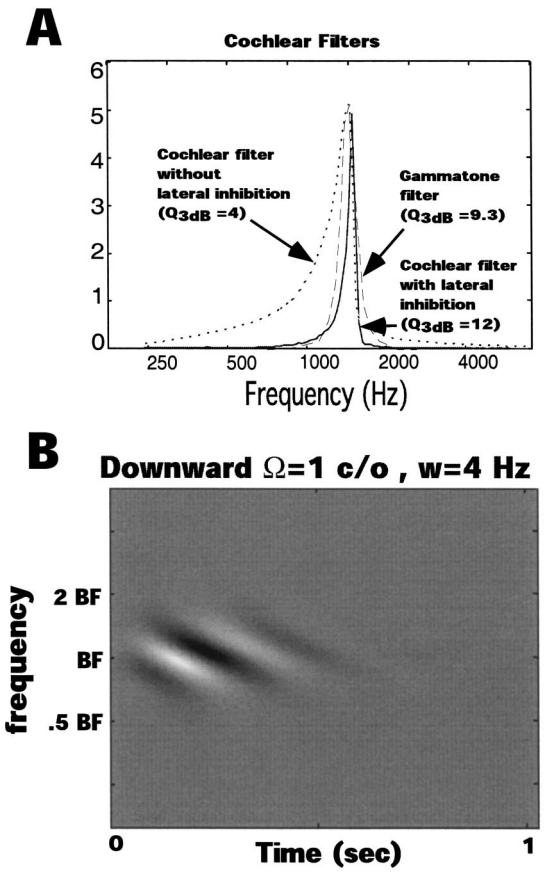


FIG. 2. (a) Frequency response of the cochlear filters used in the present model with (solid line) and without (dotted line) the LIN. A gammatone filter (dashed line) is shown for comparison. (b) Spectro-temporal response field of one modulation filter, tuned to a temporal rate of 4 Hz and a spectral scale of 1 cycle per octave (see text for details).

In a previous version of the model (Shamma, 1985) the

output of this second stage was subjected to a first difference operation across the channel array [mimicking the action of a lateral inhibitory network, or ''LIN'' (Shamma, 1985)].[1] The effect of this additional processing is to sharpen the final spectral representation, leading to $Q_{3\,dB}$'s of about 12, and an ERB for the 1500-Hz-CF filter shown in Fig. 2(a) (solid lines) of 92 Hz. As the LIN is not essential for the working of the overall model, and to simplify the analysis, we drop this stage from the computations of the auditory spectrograms in most of the remainder of this paper. The effects of sharpening the spectral representation by the LIN are discussed, for a subset of conditions, in Sec. IV. C.

2. The central auditory system

The second analysis stage is a linear approximation to some aspects of the action of higher central auditory stages (especially the primary auditory cortex). Specifically, this stage estimates the spectral and temporal modulation content of the auditory spectrogram. It does so computationally via a bank of modulation selective filters that are tuned to a range of temporal modulation *rates* (ω) and spectral resolutions or *scales* (Ω), and are centered at different frequencies along the tonotopic axis (Chi *et al.*, 1999). Each modulation filter has a spectro-temporal impulse response (usually called spectro-temporal response field—STRF). An example of the STRF of one of the many filters in the bank is shown in Fig. 2(b). Three features of this particular STRF can be seen from the plot: (*i*) It is centered on a particular best frequency (BF). The location of the excitatory (black or dark) and inhibitory (white) stripes on the vertical axis indicates that it is sensitive to frequencies of about a 2-oct range around the BF (between about 0.5 and 2 BF). (*ii*) The distance between the excitatory (black or darker) portions on the time axis corresponds to a rate sensitivity of about 4 Hz. (*iii*) The excitatory portions are separated on the vertical axis by about 1 oct, giving rise to a spectral ''scale'' sensitivity of about 1 cycle per octave. Finally, the bars sweep downwards diagonally from the top left, which is denoted in the model by assigning a positive sign to the rate parameter; bars sweeping up from bottom left to top right are designated by negative rate values. This distinction reflects the differential sensitivity of neurons in the auditory cortex to the direction in which spectral peaks move (Depireux *et al.*, 2001).

The range of parameter values used in the model implementation are based on data from cortical physiology and human psychoacoustics employing spectrally and temporally modulated stimuli (Kowalski *et al.*, 1996). Specifically, we shall assume that human subjects are primarily sensitive to scales of 0.25–8 cycles/octave, rates of 2–512 Hz, and frequencies up to 8 kHz. The bandwidth and other parameters of the STRFs are motivated by physiological findings in the cortex (Shamma *et al.*, 1995; Kowalski *et al.*, 1996). For instance, STRFs typically have a scale and rate tuning $Q_{3\,dB}$ of about 1 around their center (best) scale and rate. The overall sensitivity of the total auditory model to spectral and temporal modulations is described in detail in Chi *et al.* (1999). It is, however, worth noting two further facts: (*i*) the tuning of the STRF in Fig. 2 along the time and frequency axes can be represented by a single excitatory peak, flanked by two inhibitory ''troughs,'' and (*ii*) sensitivity to different scales and rates can be obtained by compressing or expanding this function on either the frequency or time axis. Sensitivity to different CFs is obtained by *shifting* the STRF on the (logarithmic) frequency axis.

The filter output is computed by a convolution of its STRF with the input auditory spectrogram. Consequently, the filter output is large only if the spectro-temporal modulations in the auditory spectrogram are commensurate with the rate, scale, and direction of the STRF; i.e., each filter will respond best to a narrow range of these modulations. The output of the model consists of a map of the responses across the filterbank, with different stimuli being differentiated by which filters they maximally activate. The response map provides a unique characterization of the spectrogram, one that is sensitive to the spectral shape and dynamics over the entire stimulus.

The operations used to compute the model outputs are detailed in the Appendix. Here, we simply note that the final output ($\mathbf{r}(.)$) is a function of time (t), frequency or CF (x), scale (Ω), and rate (ω)—($\mathbf{r}(\mathbf{\Omega}, \boldsymbol{\omega}, \mathbf{x}, \mathbf{t})$), To display this function we will usually collapse the time axis since all stimuli considered in this paper are stationary and consequently the response magnitude becomes steady following the stimulus onset. We then display two-dimensional patterns by collapsing across one of the remaining dimensions (frequency, scale, and rate—x, Ω,ω). The response to Carlyon and Shackleton's (1994) ''synchronous harmonics'' stimulus [Fig. 1(a)] is shown in the upper row of panels in Fig. 3(a). This stimulus has most of its energy in frequency regions around 1375–1875 Hz and 3900–5400 Hz. Consequently, those STRFs that are maximally activated are tuned close to these frequencies, as is most easily seen by looking at the distribution of activity along the abscissa of the CF-scale plot shown in the middle panel. The two bands of maximum energy are separated by approximately 1.5 oct, and so maximally activated STRFs are (broadly) tuned to scales (Ω) of less than 1 cycle/octave (the mean is 0.65 cycles/octave)—as seen in the scale-rate plot in the right-hand panel. The activation also extends down to very low (coarse) scales. This occurs because the STRFs tuned to very low scales each contain an excitatory peak on the frequency axis that is sufficiently broad to encompass both spectral peaks in the stimulus.

The top right-hand panel of Fig. 2(a) also reveals that STRFs tuned to ''positive'' rates respond more strongly than those tuned to ''negative'' rates. This results from the fact that responses of filters to our pulse trains occurs very slightly later with decreasing CF (due to the narrower bandwidth of those filters), so that a close-up of the auditory spectrogram in Fig. 1(a) would reveal that dark vertical bars in each frequency region actually slope slightly down and to the right. Finally, the pulses in each frequency channel repeat at a rate of 88 Hz, and so the STRFs maximally responsive to the stimulus are tuned to rates (ω) close to 88 Hz (see abscissa of right-most panel). The response to the ''asynchronous harmonics'' stimulus (second row of panels) differs from that to the synchronous stimulus in the scale-rate plot (rightmost panels) where the maximal activation is now con-

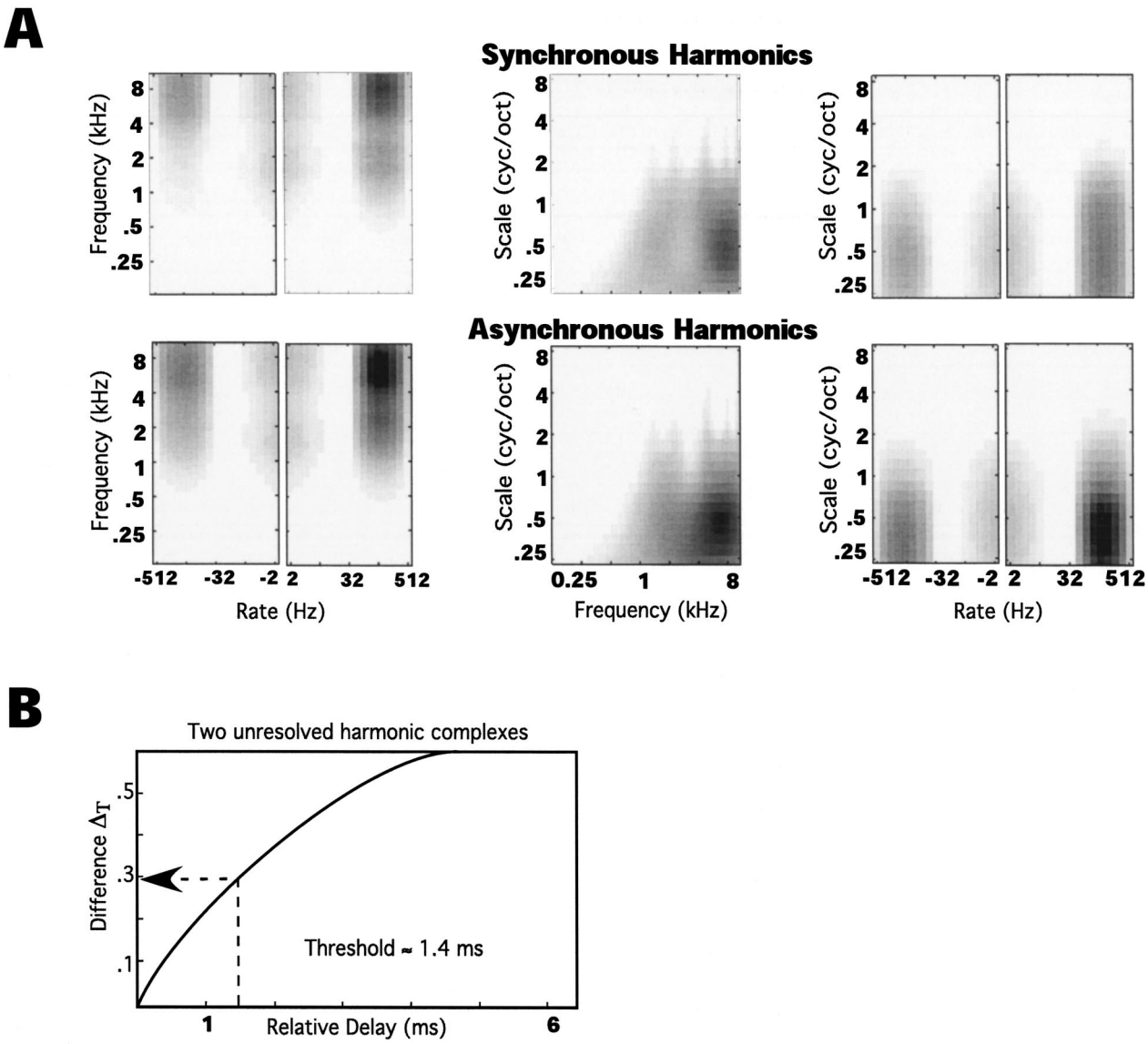


FIG. 3. Model responses to the unresolved harmonic stimuli. (a) A series of panels displaying the cortical model outputs to the two unresolved harmonic stimuli. The upper series shows the responses to the in-phase stimulus; the lower panels are in response to a 5-ms relatively delayed stimulus. Note that the biggest difference occurs between the pair of scale-rate plots (right panels). The smallest is between the rate-frequency plots (left panels). Although the STRFs were calculated at discrete values of the scale and rate parameters (with adjacent filters separated by a factor of 2), our *plots* are continuous, with the outputs of intermediate filters calculated by interpolation. (b) The normalized maximum difference between the total cortical output Δ_T is plotted as a function of the relative delay between the harmonic tone groups. Threshold is nominally set at $\Delta_T = 0.3$ corresponding to the experimentally observed relative delay at threshold of 1.4 ms.

centrated at lower scales (<0.5 cycle/octave; note that the vertical ''band'' of activation on the far right of each plot is shorter and darker in the lower panel). This is due to the cross sections in the auditory spectrograms having mostly single peaks for the asynchronous stimuli [cf. Fig. 1(b)].

3. Computing the detection thresholds

As a stimulus parameter, such as the relative delay between the two groups of harmonics in Fig. 1, is varied, a corresponding change occurs in its spectrogram and in the final output of the model (e.g., as in Fig. 3). We shall assume that human subjects can detect a certain percentage change (Δ_T) in the model output. Therefore, the detection threshold of a change in any stimulus parameter will be taken to be that which produces a Δ_T change in the model output defined as the normalized maximum change (along any axis):

$$\Delta_T = \text{maximum}|r_s - r_r|_{\Omega,x,\omega} / \text{maximum}|r_r|_{\Omega,x,\omega}, \qquad (1)$$

where r_s and r_r are the model responses to the signal and reference (standard) signals, and Δ_T (''delta threshold'') is the fractional change in the maximum value of the output.[2] We initially will estimate this value using the known results of the experiment performed by Carlyon and Shackleton (1994) and described in Fig. 3. For these stimuli, the value of Δ_T is dominated by the scale parameter Ω, which corresponds to about 0.65 cycles/octave when the pitch pulses in the two regions are in phase (separation between pulses at any one time = 1.5 oct, Ω is reciprocal of this), and to a lower range of values when the pulses are out of phase (as

explained above). Carlyon and Shackleton showed that subjects detected the difference between the in-phase and the asynchronous stimuli when the delay was approximately 1.4 ms, which corresponds to $\Delta_T = 0.3$ [Fig. 3(b)]. Therefore, we shall take $\mathbf{\Delta_T = 0.3}$ as the perceptual threshold for all remaining tests, and compute accordingly the detection thresholds for various parameter changes.

4. Psychoacoustical results to be accounted for by the model

The model above will be used in the following section, without modification, to account for discrimination or phase sensitivity in four additional paradigms: (*i*) discrimination of phase differences between pairs of harmonically related tones, (*ii*) detection of relative phase shifts between AM tones at various modulation rates and carrier tone separations, (*iii*) discrimination between AM and QFM tones at various modulation rates and carrier tone separations, and (*iv*) discrimination of stimuli having phase transitions in different frequency regions (''Huffman sequences''), as a function of the bandwidth of and frequencies at which these phase transitions occur.

III. RESULTS

A. Two resolved partials

In this section we demonstrate why the model correctly predicts that subjects cannot, in the absence of within-channel interactions and distortion products, detect phase differences between resolved partials of a complex tone. To do so we use a very simple stimulus, consisting of a dyad whose frequencies are 200 and 400 Hz. Craig and Jeffress (1962) reported that phase differences between two low-frequency tones, separated by an octave, were audible when the lower tone had a level of 50 dB SPL or higher, but not when its level was 40 dB SPL. This level dependence is consistent with phase sensitivity for these stimuli being limited to conditions where the two tones interact within a single auditory filter. A related finding is that subjects can detect ''beats of mistuned consonances'' between two tones that are slightly mistuned from an integer ratio (the small mistuning being equivalent to a time-varying phase change applied to one tone). Again, this is possible only at high overall levels, and has been attributed to interactions between the two carriers on the basilar membrane (Plomp, 1967).[3]

The output of the cochlear stage of the model to a 200 + 400 Hz dyad is shown in Fig. 4(a) for the case where the components are summed in phase, and in Fig. 4(b) for the case where they are summed 180° out of phase. The change in the model output (Δ_T) between the two phase relationships shown in Figs. 4(a) and (b) was only 0.2. Recall that a (just detectable) 1.4 ms asynchrony in the stimuli described by Carlyon and Shackleton (1994) produced a larger change of $\Delta_T = 0.3$. For those stimuli, the model was sensitive to the asynchrony because, when the two groups of harmonics moved out of alignment, there was a reduction in the output of STRFs tuned to scales of around 0.65 cycles/octave (Fig. 1). For the harmonic dyad shown in Fig. 4, however, no such change occurs. One reason for this is evident from the auditory spectrograms shown in Figs. 4(a) and (b). The phase of the response to each component varies markedly over a small range of characteristic frequencies (CFs) of 200 and 400 Hz; such transitions near the peak of the traveling wave have been observed in measurements of the basilar membrane (BM) (Dallos *et al.*, 1996) and of the auditory nerve (AN) fiber array (Kim *et al.*, 1980). As a consequence, any given time slice contains peaks of activation separated by about an octave, and this is true regardless of whether the components are summed in phase or 180° out of phase. Accordingly, STRFs tuned to about 1 cycle/octave are activated in both cases, and the model predicts that the stimuli should be hard to discriminate ($\Delta_T = 0.2$). When we artificially removed these phase transitions [Figs. 4(c) and (d)] the change in model output increased to $\Delta_T = 0.33$—the reason being that every other cycle of the 400-Hz component produced a response that was roughly synchronous with that of the 200-Hz component when they were summed in phase, but not when one component was shifted by 180°.

There are two further reasons why the model is insensitive to phase differences between resolved partials of a complex tone. One of these is that, unlike the case with envelope phase differences, where the two temporal variations to be compared have the same rate, different partials by definition have different frequencies. So, whereas for the synchronous stimulus of Fig. 1(a) all ''pitch pulses'' are in alignment, this is only true for every other cycle of the phase-straightened response shown in Fig. 4(c). When we studied the model output to a 200 + 800 Hz dyad, we found that a 180° phase change on one component produced only a small change in the model output even when we artificially straightened the responses. This occurred because now only one in four cycles coincided, and so the outputs of STRFs tuned to scales of 0.5 cycles/octave (the reciprocal of the two-octave separation between 200 and 800 Hz) were only weakly activated. Finally, at moderate and high center frequencies, the temporal fluctuations in the cochlear response to a resolved partial are smoothed by the 1-ms time constant. The usefulness of fast fluctuations is further reduced by the fact that the maximum rate to which the (model) cortical STRFs are tuned is 512 Hz.

B. AM versus QFM tones

One of the first demonstrations of listeners' sensitivity to phase differences was provided by Matthes and Miller (1947). They described the differences in the percept of an amplitude modulated (AM) tone produced by shifting the phase of the center component by 90°. This manipulation results in a ''quasi-frequency modulated (QFM)'' tone. They stated that, for carrier frequencies f_c of 500, 1000, and 2000 Hz, and a tone level of 60 dB SL, the phase difference was audible, provided that the modulation frequency (f_m) was less than 40% of the carrier frequency. A more extensive study was performed by Goldstein (1966), who measured the discrimination between AM and QFM for carrier frequencies ranging from 250 to 16 000 Hz, levels between 20 and 60 dB SL, and a stimulus duration of 1 s. The AM and QFM were presented in an alternating pattern, separated by 500 ms, and,

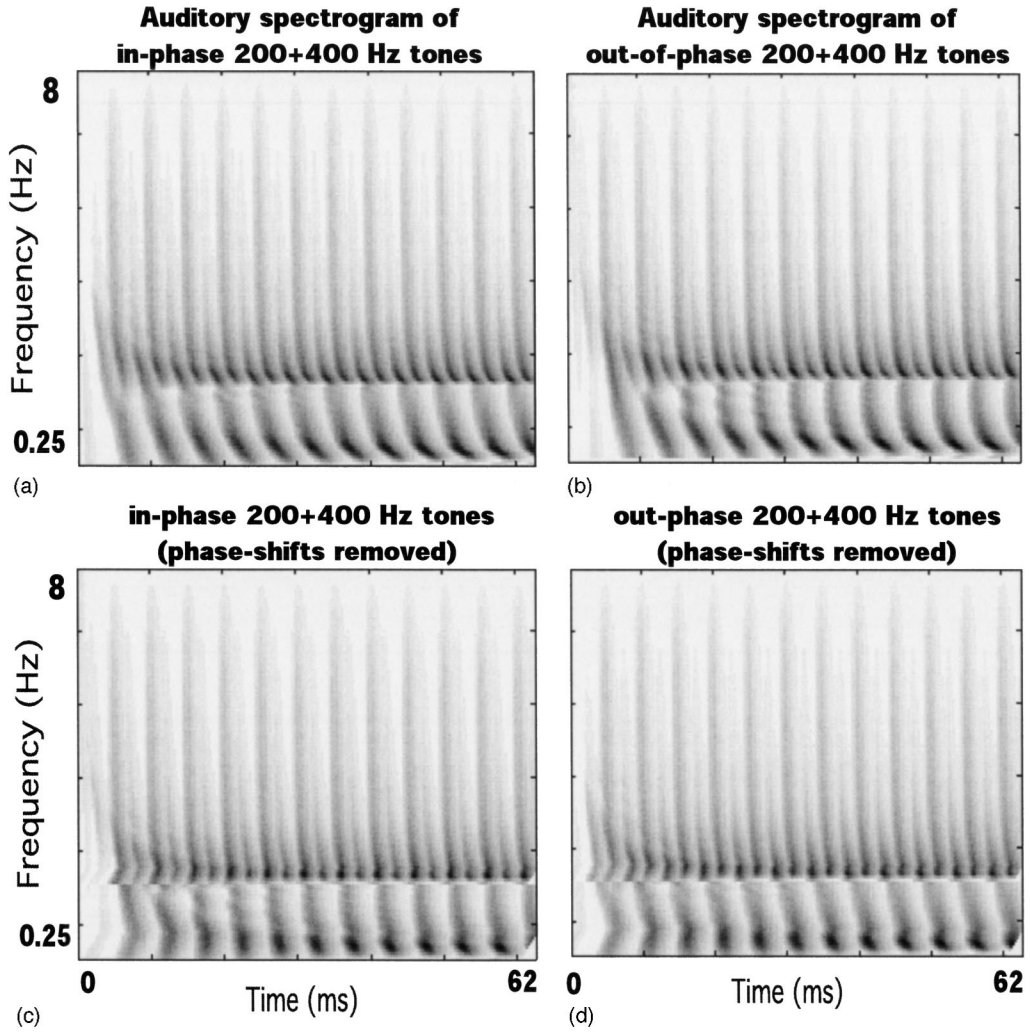


FIG. 4. Model responses to relatively phase-shifted single tones. Parts (a) and (b) show that spectrograms of two-tone stimuli (200+400 Hz) are nearly indistinguishable despite a 180° phase-shift of the 200-Hz tone between the two panels. When the two spectrograms were presented to the model described in Sec. II B., the difference between them was below the threshold for discriminability adopted by that model. Parts (c) and (d) are similar to (a) and (b), except that the phase transitions around 200 and 400 Hz have been removed by ''straightening'' the auditory spectrograms. When this is done, the difference between the resulting spectrograms is, according to the model, above threshold.

for each modulation rate, the subject simply reported whether the two sounds were the same or different. At a signal level of 60 dB SL, a difference could be heard when the ratio f_m/f_c was 0.3 or less.

Figure 5(a) displays the auditory spectrograms of the AM and QFM stimuli with a 75-Hz modulation rate and a carrier frequency of 1000 Hz. It is evident that the two spectrograms are similar except for a deeper temporal modulation of the AM tone, and a different phase structure near the CF of the tones. These differences are captured in the scale-rate plots [Fig. 5(a)—right panels], by a marked decrease in the output of the QFM at $\omega = \pm 75$ Hz. As modulation rate increases, the response of the cochlear filters to the AM tones shows a smaller modulation depth, due to the spectral sidebands being attenuated by the filter skirts. This causes the difference in the model output (Δ_T) between the AM and QFM tones to decrease at higher modulation rates; as shown in Fig. 5(b), it becomes perceptually undetectable around 330 Hz. This is consistent with the thresholds of 400 and 300 Hz, reported by Mathes and Miller (1947) and by Goldstein (1966), respectively. Figure 5(b) also shows that thresholds vary monotonically with carrier frequency (f_c), but that the predicted increase from $f_c = 1000$ Hz to $f_c = 4000$ Hz is less than proportional. In fact, the increase corresponds to a factor of 2.3 for a fourfold change in f_c. An inspection of Goldstein's (1966) data also shows a less than proportional increase, although this corresponds to a factor of about 3, which is slightly greater than seen here.

The reason for the less than proportional increase in predicted thresholds at the two highest values of f_c shown, 2 and 4 kHz, occurs because these thresholds are then 512 Hz, which is the highest temporal modulation rate to which an STRF is tuned. The result is that the modulation depth in the auditory spectrogram has to be increased by reducing the modulation rate. The corresponding finding reported by Goldstein (1966) may have a similar basis.

C. AM versus phase-shifted AM tones

Yost and Sheft (1989) required listeners to discriminate between two stimuli, each consisting of a pair of sinusoidally amplitude modulated tones, differing only in the relative

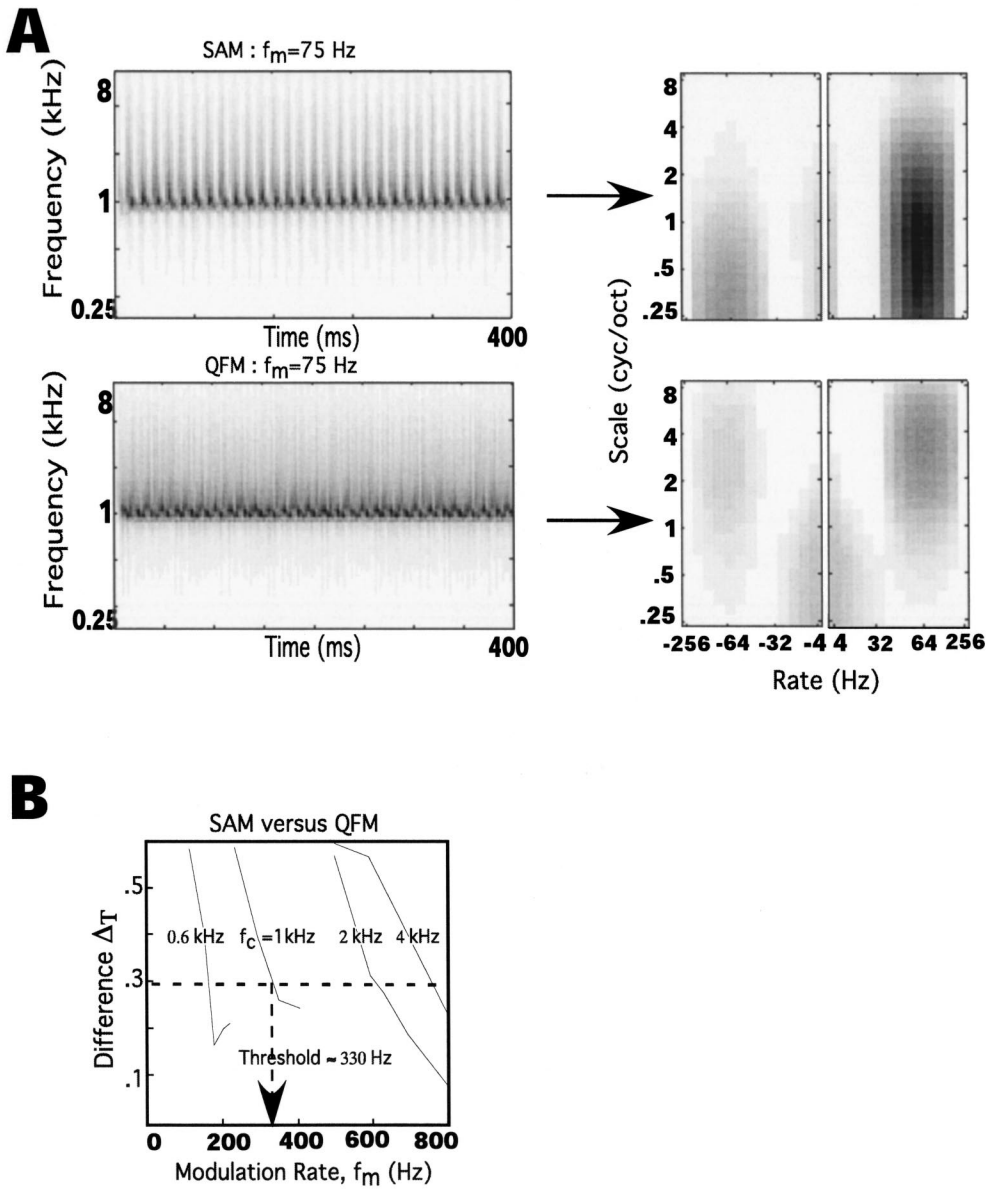


FIG. 5. Model responses to sinusoidally amplitude modulated (SAM) and a quasi-frequency modulated (QFM) tone. (a) Spectrograms and model cortical outputs in response to a 75-Hz SAM tone (top panels) and a 75-Hz QFM tone (bottom panels). Only the scale-rate plots of the model cortical outputs are shown since this is the plane with largest differences between the two stimuli. (b) The normalized maximum difference between the total cortical output Δ_T is plotted as a function of the modulation rate. Threshold is predicted to be at a modulation rate where $\Delta_T = 0.3$, i.e., it is approximately 330 Hz for a carrier frequency of 1 kHz; predictions for other carrier frequencies are also shown.

phase of the modulators applied to each carrier. In each 2AFC trial, the modulators were in phase for one stimulus (the ''standard''), and delayed on the lower carrier for the other (the ''signal''). The stimuli were presented against a continuous background noise, and thresholds were defined as the phase delay leading to 75% correct, as estimated from a two-point psychometric function. When the two carrier frequencies were 1000 and 4000 Hz, threshold corresponded to a modulator delay of about 60°. This threshold was roughly independent of modulator frequency over the range studied (5–20 Hz). Strickland *et al.* (1989) obtained a broadly similar threshold using the same carrier frequencies, but with an adaptive procedure and an 8-Hz modulator that was delayed on the upper carrier during the signal interval. Increasing the modulation rate to 128 Hz resulted in a marked deterioration in performance.

Figure 6(a) shows the auditory spectrograms of two pairs of AM tones, having carrier frequencies of 1 and 4 kHz in each case, together with the resulting scale-rate plots. When the carriers are two octaves apart, and there is no phase lag, there is a maximum activation in STRFs tuned to the reciprocal of this separation (0.5 cycles/octave) and to the modulation rate (30 Hz). There is a fair degree of spread towards STRFs tuned to lower scales; as discussed earlier, this is because a low-scale STRF will contain an excitatory peak that is broad enough to encompass both spectral peaks in the stimulus. The rate-scale plot to a stimulus with a 90° phase lag appears similar, but, as shown in Fig. 6(b), differs slightly in the output of filters tuned to low spectral scales (approximately 0.25 cycles/octave; cf. Sec. II B 2, Fig. 3).

Figure 6(c) summarizes the differences Δ_T between the two tones for different phase-shifts (both with the modula-

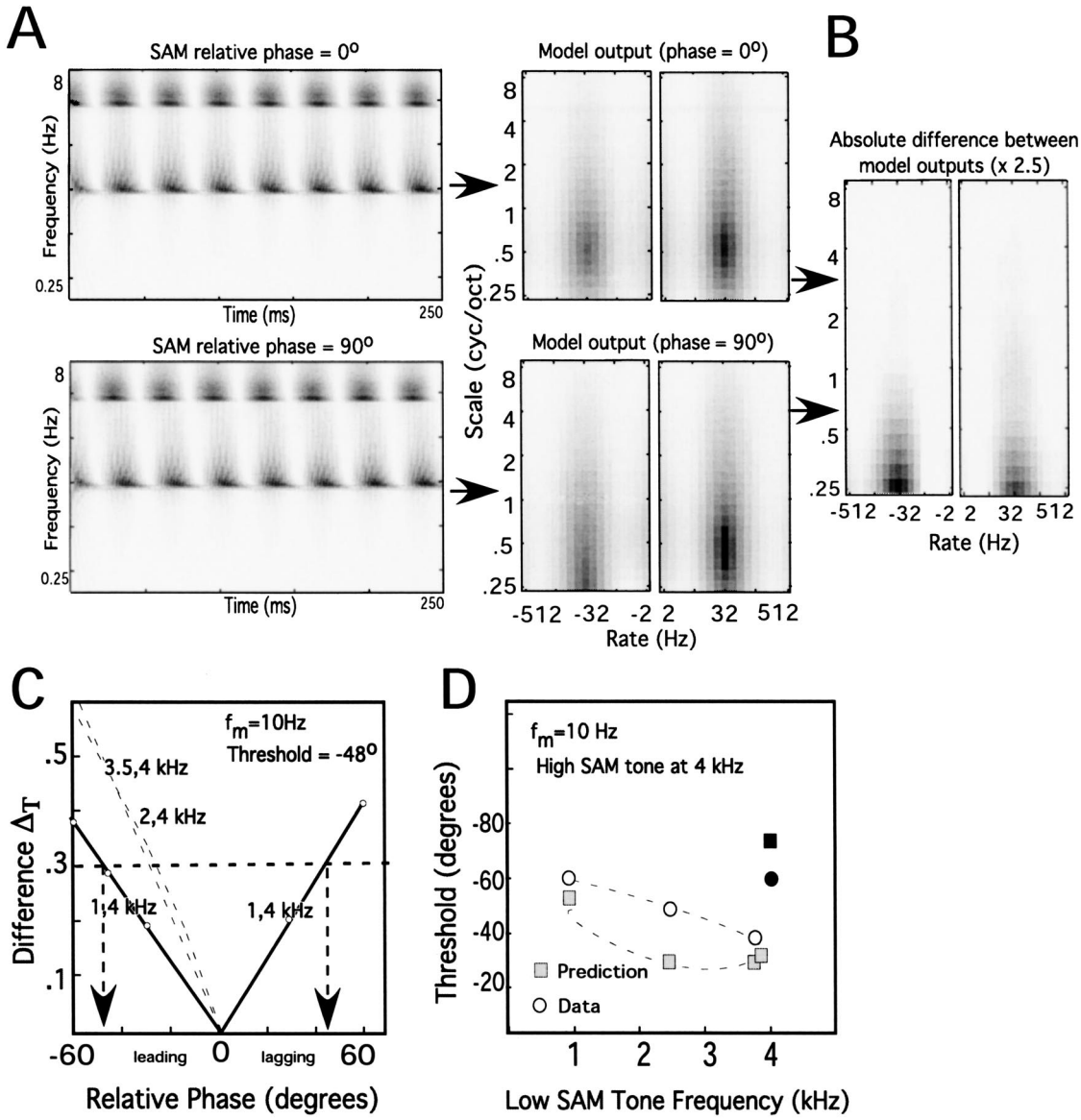


FIG. 6. Model responses to sinusoidally amplitude-modulated (SAM) tones. (a) Spectrograms and model cortical outputs due to in-phase pair of SAM tones at 1 and 4 kHz (top panels) and relatively phase-shifted SAM tones (bottom panels). Only the scale-rate plots of the model cortical outputs are shown since this is the plane with largest differences between the two stimuli. (b) Difference between the scale-rot plots to the two stimuli shown in part (a). (c) The normalized maximum difference between the total cortical output Δ_T is plotted as a function of the relative phase-shift between the SAM tones, for a modulation rate of 10 Hz. Threshold is predicted to be at $\Delta_T = 0.3$, i.e., it is approximately 48° for the lower-tone leading configuration. Thresholds are also shown for different SAM tone combinations, where they generally decrease with smaller frequency separation. (d) Thresholds as a function of the carrier frequency of the lower tone, as predicted by the model (open squares) and obtained by Yost and Sheft (open circles). The unconnected filled square shows the predicted threshold when both carriers are at 4 kHz. Data obtained by Fleischer (1980) for a condition analogous to this are shown by the filled circle (see text for details).

tion on the lower tone leading and lagging). Since the perceptually detectable difference is $\Delta_T = 0.3$, the corresponding predicted detectable phase-shift is approximately 48° for the condition with the lower tone leading (as studied by Yost and Sheft). Interestingly, the model predicts that thresholds should be slightly lower, at 42°, when the envelope of the lower tone lags. This occurs because the responses of filters tuned to lower CFs occur slightly later than those tuned to higher CFs, and because this difference is exaggerated when an additional delay is imposed on the lower tone's envelope. Unfortunately, neither Yost and Sheft nor Strickland *et al.* examined the effect of which tone was leading or lagging, and so we do not know whether this small difference would occur experimentally.

An additional effect present in the model predictions is that, for carrier frequencies of 1000 and 4000 Hz, thresholds increase by about 10° as modulation rate is increased over the range (5°–20°) studied by Yost and Sheft (1989). Those authors did not mention any such increase, but the variation predicted by the present model is quite small, and within the variability of their data. As shown by Strickland *et al.* (1989), performance deteriorates markedly when the modulation rate was increased to 128 Hz, and our model can capture this fact, leading to a predicted threshold of 130°. The

reason for this increase is that, as the side-bands produced by the AM are attenuated by the slopes of the cochlear filters, their output becomes less deeply modulated.

Another aspect of the predictions that is in general agreement with the data is the variation in threshold as the frequency of the lower component is raised from 1 to 3.5 kHz. The open symbols in Fig. 6(d) show a decrease in threshold for both the predictions (squares) and Yost and Sheft's data (circles). An interesting additional observation is that the model predicts that, for very small frequency separations, smaller than the smallest separation used by Yost and Sheft, thresholds should start to *increase* again. In the extreme case, where the two carriers have the same frequency, thresholds are predicted to rise to 75°, as shown by the unconnected solid square to the right. This increase occurs because the modulator phase difference no longer produces any change in the scale parameter, and the task becomes that of detecting a change in modulation depth. More specifically, it can be shown[4] that, when the carrier frequencies are identical, summing two tones having a modulator phase difference of θ rad produces a pure tone modulated at a depth equal to $\cos(\theta/2)$. When $\theta=0$ the depth is equal to 1, and Fleischer (1980) has shown that subjects can just discriminate between such a 100% modulated tone and one with a modulation depth of 0.89. This latter value would be produced by summing two modulated sinusoids with $\theta=0.95$ rad, or 54°, and so we would predict that, if Yost and Sheft had reduced the frequency separation between their carriers even further, their thresholds would have risen to this amount, indicated by the unconnected solid circle in Fig. 6(e).

D. Discrimination of Huffman sequences

Patterson and Green (1970) studied the discrimination of two clicklike stimuli differing only in the frequency region that contained a 360° phase transition. These ''Huffman sequences'' have identical power spectra and were generated by passing a click through a digital all-pass filter. The filter can be characterized in the z-plane by a single pole-zero pair (plus another pair at the corresponding negative frequency), with the pole and the zero equidistant from, and on opposite sides of, the unit circle. The frequency at which the transition occurs is determined by the angle of the pole-zero pair from the origin, with the distance from the unit circle being related to the bandwidth of the transition. They showed that Huffman sequences as brief as 2.5 ms could be discriminated,[5] indicating that the ear is sensitive to phase differences that produce very short differences in the time-of-arrival of energy in different frequency regions. The difference between pairs of sequences was reported as being phenomenologically akin to that between ''tick'' and ''tock,'' and they noted that subjects required considerable training and full attention in order to perform the task successfully. Sensitivity was measured using a 2IFC task and the method of constant stimuli. For a stimulus duration of 10 ms, and transitions centered on 400 and 800 Hz, performance improved from chance at a transition bandwidth of 10 Hz to more than 90% correct at a bandwidth of 80 Hz. Sensitivity differed somewhat across listeners, but the average threshold, defined here as the bandwidth corresponding to 70% correct, was about 20 Hz. Another experiment measured the improvement in discriminability between two Huffman sequences as a function of the difference between the frequencies at which the transition occurred.

Figure 7(a) displays the auditory spectrograms of two Huffman sequences that differ by the frequency at which their phase shift occurs (800 and 400 Hz). The spectrograms illustrate the phase delay experienced by the responses at the different CF locations. The difference in activation across CF is the most important feature that distinguishes these spectrograms; the phase shifts appear as discontinuities in the responses at different CFs, which are detected and represented by the model at different CF locations. This can be seen in the rate-frequency panels to the right. The bandwidth of the phase-shifted regions in this figure is 50 Hz. As the bandwidth is reduced, the difference between the two delayed regions becomes smaller, and hence less detectable. The solid line in Fig. 7(b) summarizes the results for several bandwidths. The results for a very different CF pair are shown by the dashed line. The curves predict that the two sequences become increasingly alike as bandwidth is reduced, and are indistinguishable below about 10–15 Hz, approximately as found by Patterson and Green (their Fig. 5). They noted that the amount of delayed energy introduced by the phase discontinuity decreases as bandwidth is reduced, and used this to explain why performance also became poorer at narrow bandwidths.[6]

In another test, the difference between the frequencies of the phase-shifted regions was varied while holding bandwidth constant at 50 Hz. Figure 7(c) shows that, as the frequency difference (Δf) is reduced, the responses to the two stimuli converge, reaching a threshold at around 125 Hz—close to the value of about 150 Hz observed by Patterson and Green (1970, right panel of their Fig. 6). At lower overall frequencies and the same Δf, however, the sequences become detectable again [Fig. 7(c), dashed lines], as also shown in the experimental results (Patterson and Green, 1970, left panel of their Fig. 6). A discrepancy between the model and the data is that Patterson and Green reported that, when the lower tone had a frequency of 400 Hz, performance was close to ceiling at all frequency separations studied, which ranged from 100 to 400 Hz. In contrast, the model predicts that sensitivity should decrease at small frequency separations. Hence, although our model successfully accounts for the superior performance at low compared to high carrier frequencies, and for the effect of frequency separation at high carrier frequencies, it fails to predict the ''ceiling'' effect observed at low carrier frequencies.

IV. DISCUSSION

A. Previous models of monaural phase perception

In the preceding section we presented a model that can account not only for listeners' inability to detect across-channel differences in fine structure between resolved partials, but also for their ability to detect across-channel envelope differences. We have argued that representations which discard across-channel timing information, such as the ''au-

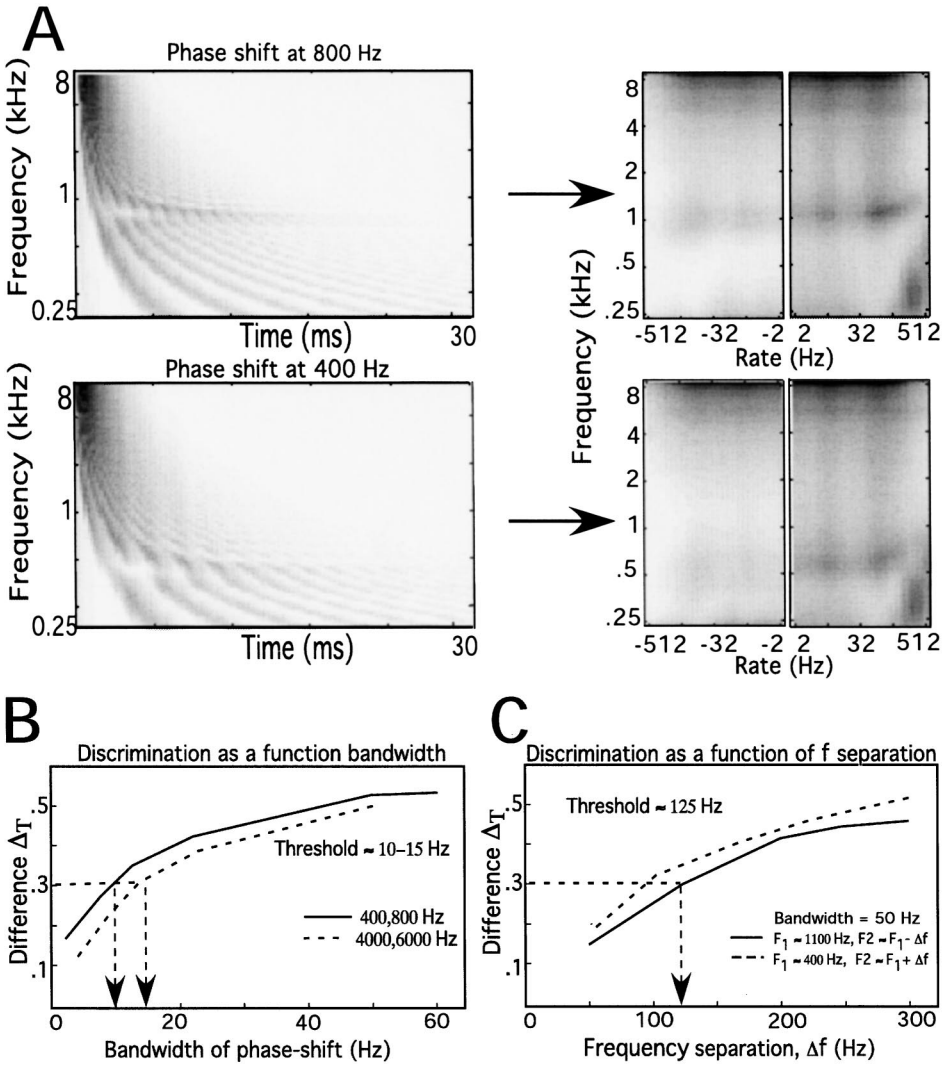


FIG. 7. Discrimination of Huffman sequences as a function of bandwidth. (a) Auditory spectrogram of two Huffman impulses with phase-shifts at 400 and 800 Hz and a bandwidth of 50 Hz (see text for details of the stimulus). The most significant difference in the model output occurs in the rate-frequency plane. Specifically, note the difference in the frequency location of the outputs. (b) As bandwidth is reduced, the clarity of the phase-shifts diminishes gradually, and the difference between the two stimuli is predicted to become perceptually undetectable at a bandwidth of 10 Hz (or 15 Hz for transitions at 4000 and 6000 Hz). (c) Difference in model output between two Huffman sequences (transition bandwidth = 50 Hz) as a function of the difference between the frequencies at which the phase transitions occur. Solid line: fixed frequency (F1) = 1100 Hz, F2 = F1 − Δf. Dashed line: F1 = 400 Hz, F2 = F1 + Δf.

ditory image'' (Patterson *et al.*, 1995) and ''autocorrelogram'' (Meddis and Hewitt, 1991a; Meddis and O'Mard, 1997) models, do not account for this latter finding. Here, we examine this claim further, and suggest ways in which the models would have to be modified in order to account for the data.

We have measured the output of Meddis and O'Mard's (1997) autocorrelogram model to two experiments successfully modeled here. [Although described in terms of pitch perception, Meddis and O'Mard's model is essentially an updated version of the Meddis and Hewitt (1991a, b) model, which was designed to account for both pitch and phase sensitivity.] The two experiments involved the discrimination of asynchronies between groups of unresolved harmonics (Carlyon and Shackleton, 1994), and between the envelopes of sinusoidally amplitude modulated tones (Yost and Sheft, 1989). In both cases, the summary autocorrelogram (SACF), when averaged along the entire stimulus, did not change systematically when an asynchrony was introduced. This finding follows naturally from the fact that the autocorrelogram in each frequency region is calculated separately, before being summed to form the SACF. It is worth noting, however, that for each stimulus, the overall height of the SACF varies markedly over time, as illustrated in Fig. 8, which shows the SACF to the same sound—the synchronous stimulus Carlyon and Shackleton (1994)—measured at two different times. This occurs because the autocorrelogram is calculated over a window that ends abruptly at a given point, and which decays into the past with a time constant of 10 ms, which is comparable to the period of the 88-Hz stimuli used by Carlyon and Shackleton, and shorter than that of the highest modulation rate (20 Hz) used by Yost and Sheft. The result is that the total power falling within the window, and hence the height of the SACF, depends crucially on the timing of the window relative to the amplitude peaks in each frequency region. One implication for the present discussion is that, because the SACF is calculated by summing the ACFs of individual channels, its height will fluctuate more when the peaks are synchronous than when there is an asynchrony. It is therefore possible that the model could be modified to account for this sensitivity, by the implementation of an extra stage that measured the amount of fluctuation in the SACF. It should be noted, though, that one could equally bypass the calculation of autocorrelograms altogether, and simply sum auditory nerve activity across channels and observer the fluctuation in that sum.

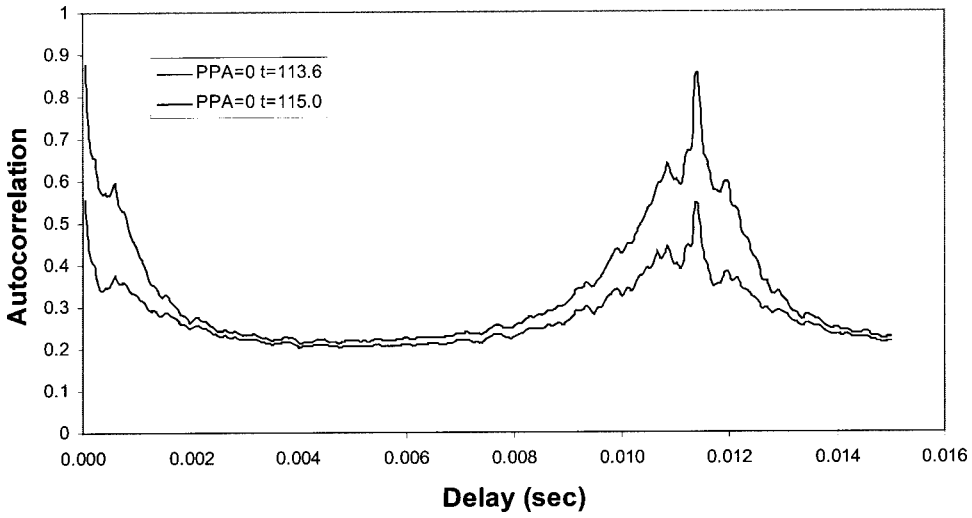


FIG. 8. The SACF obtained using Meddis and O'Mard's (1997) model in response to Carlyon and Shackleton's (1994) ''synchronous harmonic'' stimulus, measured at 113.6 and 115 ms after the stimulus onset.

Another possible solution, which applies both to the autocorrelogram and auditory image models, relies on intermediate frequency channels that respond to both of the harmonic complexes. We can illustrate this using Patterson *et al.*'s (1995) ''auditory image'' model (''AIM''), which filters the input with an array of bandpass filters, simulates peripheral processes including compression and adaptation, and sends a tonotopically arranged set of ''neural activity patterns'' to a central buffer, having a default duration of 35 ms. When the pitch pulses in two unresolved harmonic complexes are misaligned in the stimulus, they nevertheless become realigned by the model's strobing mechanism. This can be seen in Fig. 9(b), which shows the ''stabilized auditory image (SAI)'' of the pair of unresolved complexes studied by Carlyon and Shackleton (1994), with a PPA of 5 ms. Note that the pitch pulses in the regions of 1375–1875 Hz and 3900–5400 Hz have been aligned, and so the SAI appears essentially identical to the case where there is no PPA [Fig. 9(a)]. A small difference can be seen in channels having CFs between the two complexes (indicated by arrows), which show twice as many maxima for the stimulus with the PPA [Fig. 8(b)] than for the one without [Fig. 8(a)]. This interaction, which occurs because these channels respond to maxima in both pulse trains, is shown in more detail for a smaller range of CFs in Fig. 8(c) (see arrow). The output level of these intermediate channels will depend on the degree of peripheral compression used by the model [we have used $y(t)=x(t)^{0.5}$, which is often used as a ''default'' by AIM]. However, whatever the compression used, this output will be swamped by the addition of background noise; this is illustrated in Fig. 8(d), with noise added at the same signal-to-noise ratio (SNR) as in the Carlyon and Shackleton study. The figure shows that the noise has ''filled in'' the valleys between peaks in the intermediate channels (CFs around 2500 Hz, indicated by arrows), but not in those channels tuned to either of the two complexes [top and bottom traces in Fig. 8(d)]. Such within-channel cues have a much smaller effect in our present model, because it does not discard across-channel timing information and so can compare peaks of activity in channels tuned to each group of harmonics. When the noise used in the plots shown in Fig. 8(d) was added to our own simulations, the distance metric for a 5-ms PPA was reduced only slightly, from 0.6 to 0.52—still well above the threshold value of 0.3 (defined in Sec. II B 3).

Finally, we should note that, although the output of the final, strobed temporal integration (''STI'') stage of AIM cannot account for the aspects of across-channel phase sensitivity described here, it is possible that another ''route'' could be added to the model. For example, some aspects of auditory perception could still depend on STI, whereas other, perhaps parallel, processes could mediate phase and other discriminations. What is clear, however, is that AIM would require significant modifications—to the extent of requiring different representations for different tasks—for this to be the case.

B. Modifications to the computational model

Although the present model can quantitatively account for a number of findings derived from a range of different paradigms, the specific mechanisms and parameters of the model are unlikely to be crucial in this regard. The important aspects of the model are that the peripheral filters used produce significant phase transitions around peaks in the traveling wave, that responses to resolved partials are temporally smoothed, and that the central analysis stage does not discard across-channel information. One modification would be to replace the auditory spectrogram with one based on time-varying excitation patterns derived from other commonly used cochlear filter-banks (Slaney, 1993). Indeed, phase transitions similar to those shown in Fig. 4 occur not only with the filter-bank used here, but also when the filters are sharpened by a lateral inhibitory network (Shamma, 1985) or replaced by the more symmetrical gammatone filter (Patterson *et al.*, 1988). The main effect of varying the bandwidths of the filters is to influence the precise way in which the model output varies with the frequency separation between carrier frequencies, as will be discussed in more detail in Sec. IV. C. It is also true that the multi-scale modulation analysis could probably be replaced with related methods used by other authors. For example, the rate tuning is reminiscent of the ''modulation filterbank'' proposed by Dau *et al.* (1996), and

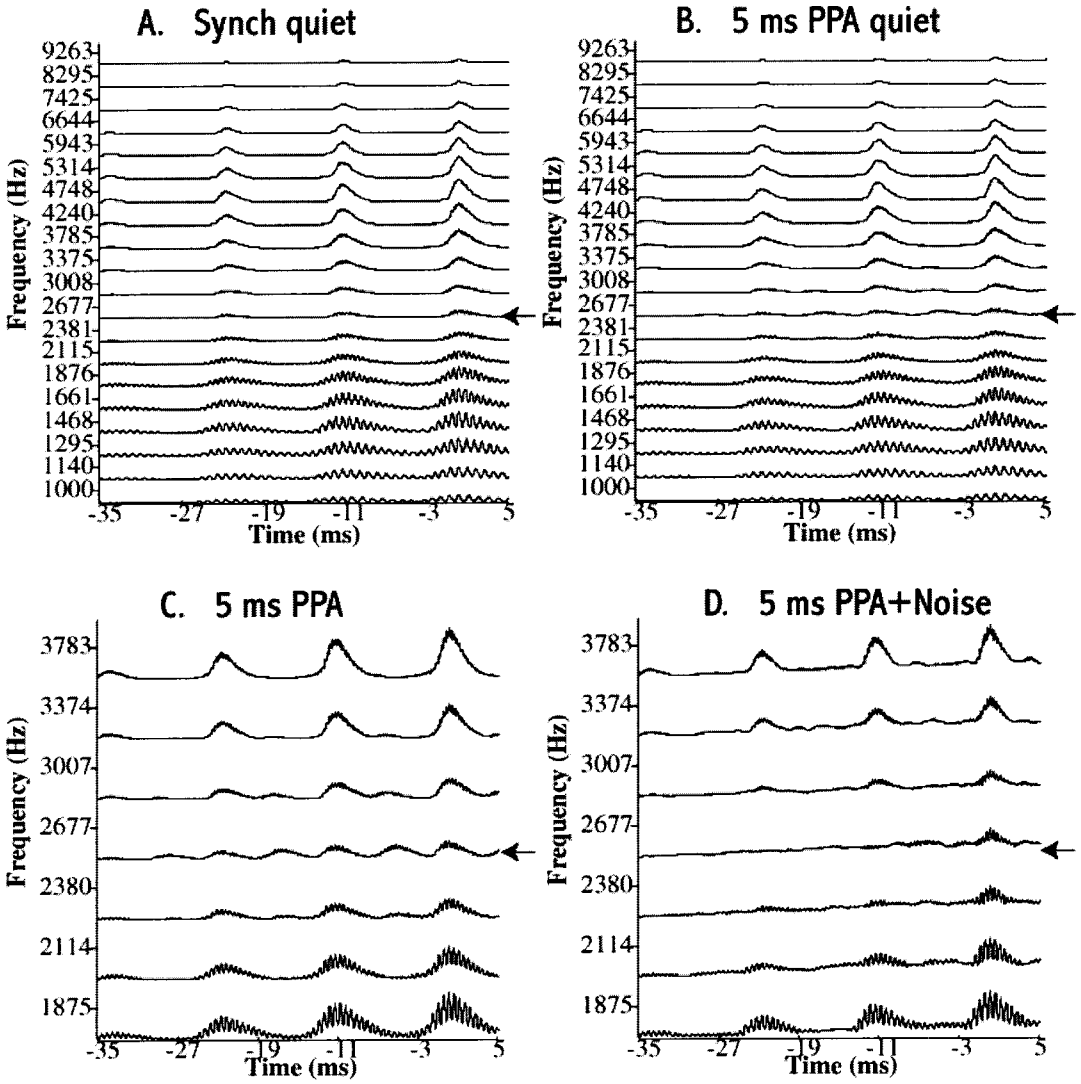


FIG. 9. (a) ''Stabilized auditory image'' generated by Patterson *et al.*'s (1995) auditory image model to the stimuli similar to those used by Carlyon and Shackleton (1994), for a stimulus in which the pitch pulses in the two frequency regions are in synchrony. The background noise used by Carlyon and Shackleton (1994) was not included in this simulation. (b) As in (a) but with a PPA of 5 ms. Part (c) is an expanded version of part (b), showing only seven frequency channels. Part (d) is similar, but with pink noise added. The spectrum level of this noise at 2 kHz was approximately 33 dB below the level of the components of each complex in the filter passband. This was the same as used by Carlyon and Shackleton.

the scale parameter resembles, in some respects, the ''Mellin transform'' recently applied to central auditory representations by Irino and Patterson (2002).

Finally, we should stress that our central mechanism was selected for physiological validity rather than simplicity, and that computationally simpler algorithms may do a good job of accounting for the data. For example, we tried to model the experimental results by simply calculating the normalized Euclidean distance between pairs of auditory spectrograms, and could successfully account for the dependence of AM vs QFM discrimination thresholds on modulation rate. However, this approach erroneously predicted that subjects should be able to detect a 180° phase shift on the lower component of a 200+400-Hz dyad, and could not account for the experimentally observed reduction in SAM phase discrimination thresholds with reducing frequency separation. More importantly, such an approach requires that subjects somehow align the start points of two auditory spectrograms to be compared, and do so with millisecond accuracy despite the stimuli being presented sequentially. It therefore seems clear that the auditory system must calculate *some* summary measure of the spectrogram, and that this measure must be sufficiently stable to allow a comparison between sequential stimuli. Our model fulfills this criterion, without sacrificing sensitivity to relative phase delays between concurrently stimulated frequency channels.

C. Effects of nonlinearity and filter bandwidths

Although the model described here does a fairly good job of accounting for data from a number of different paradigms, it does suffer from two limitations arising from the cochlear stage of the model. First, the generation of the auditory spectrogram is an entirely linear process. This simplifies the computations, and eliminates the need to specify the level of the stimuli at the input to the model, but is almost certainly wrong. It is well known that the response of the basilar membrane is highly compressive at moderate-to-high

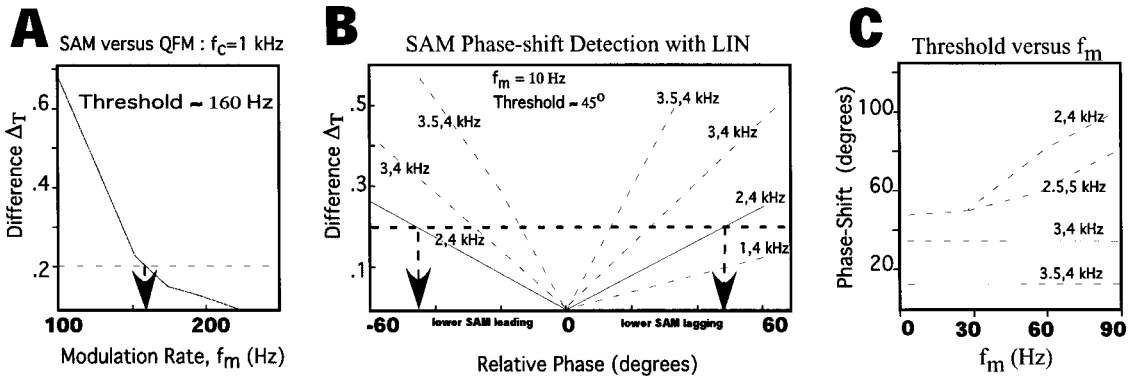


FIG. 10. Output of revised model in which the LIN has been reinstated. (a) AM vs. QFM [cf. Fig. 5(b)]. (b) Modulator phase differences between pairs of SAM tones [cf. Fig. 6(c)]. (c) Variation in predicted thresholds with modulation rate for pairs of SAM tones.

levels (Sellick *et al.*, 1982), and that the form of this compression results in a broadening of peripheral filters with increasing level (e.g., Rosen and Baker, 1994). This in turn is likely to affect performance on phase discrimination tasks that rely on within-channel interactions. For example, Goldstein (1966) reported that, for a carrier frequency of 1 kHz, the highest modulator rate at which AM and QFM could be distinguished was 300 Hz at a signal level of 60 dB SL, but 200 Hz at 20 dB SL. It is likely that the present model would have to include level-dependent filter widths in order to account for this trend, which was also observed by Mathes and Miller (1947).[7] A related point is that the cochlear filters used here, in the absence of the lateral inhibitory network (LIN) of the ''full model,'' are broader than those measured psychophysically at moderate levels. For example, the ERB of our filters is about 170 Hz at a CF of 1 kHz, compared to the 132 Hz measured from notched-noise data (Glasberg and Moore, 1990). One would expect this to affect the frequency separations at which phase differences between partials become distinguishable.

To test the influence of spectral selectivity on the thresholds predicted by the model, we have examined the effect of reintroducing the LIN, which effectively sharpens the filters by a factor of about 3, making them *more* selective than the gammatone [Fig. 2(a)].[8] To do this, we recomputed the value of Δ_T corresponding to the threshold value of PPA = 1.4 ms in Carlyon and Shackleton's data, now with the LIN reinstalled, and used this value as the new criterion for predicting thresholds in the other paradigms described here. An example is shown in Fig. 10(a), which describes the effect of modulator rate on the discrimination between AM and QFM. A comparison with Fig. 5(b) shows that the threshold has dropped from its previous value of 330 to 160 Hz, consistent with the increased spectral selectivity. Predictions for the detection of envelope phase differences between SAM tones are shown in Fig. 9(b). It can be seen that the predictions for carrier frequencies of 2 and 4 kHz are now more similar to those previously obtained for carriers of 1 and 4 kHz, [Fig. 6(c)] again consistent with the internal spectral representation having been ''sharpened up.'' Indeed, as Table I shows, whenever the results of our simulations can be expressed in terms of frequency separations or bandwidths, increasing spectral selectivity shifts thresholds to smaller values. It is, however, worth noting that the proportional change in these values is substantially less than the threefold increase in $Q_{3\,dB}$ produced by reintroducing the LIN.

V. SUMMARY

We have described a model of monaural phase perception that accounts for three facts: (*i*) listeners can detect phase differences between sinusoidal components that interact in the auditory periphery, (*ii*) they cannot do so when those components are sufficiently resolved in frequency to preclude such within-channel interactions, but that (*iii*) they *can* make quite fine (<2 ms) across-channel timing discriminations between broader-band stimuli. Whereas previous models (Meddis and Hewitt, 1991b; Patterson *et al.*, 1995; Meddis and O'Mard, 1997) successfully accounted for the first two findings, they fail to do so for the third.

Our approach consists of two stages: the calculation of an ''auditory spectrogram,'' based on well-established properties of the cochlea, and processing of the auditory spectrogram by a more central stage. The central processing we have implemented could no doubt be replaced by a number of other metrics, and still lead to qualitative agreement with the experimental findings modeled here. However, it seems important to invoke processes that are physiologically plausible, and our specific approach has the advantage of being based on physiological measurements obtained in auditory cortex. Furthermore, it includes a decision metric that produces predictions that are in reasonably good quantitative agreement with data obtained from four different paradigms,

TABLE I. Effect of reintroducing the lateral inhibitory network (LIN) on thresholds predicted by the model, for four different paradigms. Predictions for SAM tones are expressed as the frequency separation leading to a threshold of 30°, where the upper component is fixed at 4 kHz. Predictions for the AM-QFM task are for a carrier frequency of 1 kHz. The predictions for the Huffman sequences labeled ''bandwidth'' are for the stimuli corresponding to Figs. 7(a) and (b) with phase shifts at 400 and 800 Hz. Predictions labeled ''Huffman (Δf)'' correspond to those in Fig. 7(c), where the upper phase transition is at 1100 Hz.

Paradigm	No LIN (Hz)	LIN (Hz)
SAM tones	1700	1000
AM vs QFM	330	160
Huffman (bandwidth)	12	8
Huffman (Δf)	125	85

using a single set of parameters. As such it provides an existence proof that monaural phase sensitivity can be modeled without explicit removal of across-channel timing cues.

ACKNOWLEDGMENTS

The collaboration leading to this work resulted from an EPSRC (U.K.) network grant to C. J. Darwin. The second author's contribution has been supported in part by a grant from the Office of Naval Research under the MURI97 Program, to the Center for Auditory and Acoustic Research. We thank John Deeks for help with some of the simulations, Christophe Micheyl for comments on a previous version of this article, and Brian Moore for handy tips.

APPENDIX: STRF CALCULATION

For the cortical output (Fig. 2), the STRF is denoted by

$$\mathrm{STRF} = g_{RF}(t;\omega_c,\theta_c).h_{RF}(x;\Omega_c,\Phi_c),$$

where g_{RF} and h_{RF} represent the temporal and spectral impulse response dimensions of the STRF, each parametrized by a center scale and rate, and their respective phases:

$$g_{RF}(t;\omega_c,\theta_c) = g(t;\omega_c)\cos\theta_c + \underline{\underline{g}}(t;\omega_c)\sin\theta_c,$$

$$h_{RF}(x;\Omega_c,\Phi_c) = h(x;\Omega_c)\cos\Phi_c + \underline{\underline{h}}(x;\Omega_c)\sin\Phi_c,$$

where $\underline{\underline{g}}$ and $\underline{\underline{h}}$ are the Hilbert transforms of g and h. The final output $r(\cdot)$ is then computed from a convolution of the STRF and the auditory spectrogram $y(\cdot)$,

$$r(t,x;\omega_c,\theta_c,\Omega_c) = y(t,x) * xt \quad \mathrm{STRF}.$$

The output is a complex quantity that has a magnitude (envelope) and phase. For more details, please refer to the model description in Chi *et al.* (1999) [and to Wang and Shamma (1995) for the static cortical model].

[1]An earlier version of the model also included a time constant of 8 ms when implementing the LIN. This time constant is not used in the present paper.

[2]Note that, as with other models of phase sensitivity, the model does not explicitly include the effects of the variability of this internal representation, as would happen in a real biological system. By specifying threshold as a change in the model output as a proportion of the maximum overall activity, we are implicitly assuming that the standard deviation of the representation response increases as a function of overall activity.

[3]An exception may occur when binaural cues are available. Feeney (1997) reported that four out of six subjects could hear beats of mistuned consonances when the two carriers (f1 and f2) were presented to opposite ears. His subjects reported using a changing location as the cue, and he interpreted his results in terms of a specifically binaural mechanism.

[4]The waveform of two AM tones whose envelopes differ by θ rad is $y(t) = [(1+\cos(f_m)) + (1+\cos(f_m+\theta))]\sin(f_c) = [2+\cos(f_m)+\cos(f_m+\theta)] \times \sin(f_c)$. From trigonometry, this equals $2[1+\cos(\theta/2)\times\cos(f_m+\theta/2)] \times \sin(f_c)$.

[5]The most obvious way to limit the duration of a Huffman sequence to x ms is simply to delete portions of the waveform occurring after that time. We used this method in the simulations described here. Patterson and Green adopted a different technique, which was to add a second pulse whose timing, level, and polarity caused it to cancel out the waveform produced by the second pulse after x ms.

[6]Patterson and Green pointed out that the time delay corresponding to the 2π phase shifts present in Huffman sequences is equal to 1/2 BW, where BW is the transition bandwidth in Hz. This means that limiting the duration to 10 ms will have the effect of removing some of the delayed energy whenever BW$<$50 Hz. This is another reason why reducing BW may have led to an increase in thresholds.

[7]The reader may be interested to know that Goldstein argued that the level dependence that he observed "should discourage the use of these data as a definitive measure of the critical band"—presumably because, at that time, the critical band was believed to be level independent!

[8]In a linear model of auditory processing such as the one used in this paper, spectral sharpening by LIN is exactly equivalent to sharpening the cochlear filters. The situation is rather different if a compressive nonlinearity (such as the hair cell's) is sandwiched in between the cochlea and the LIN [see details in Lyon and Shamma (1996)].

Buunen, T. J. F., and Bilsen, F. A. (**1974**). "Subjective phase effects and combination tones," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, Berlin), pp. 344–352.

Carlyon, R. P., (**1994**). "Detecting pitch-pulse asynchronies and differences in fundamental frequency," J. Acoust. Soc. Am. **95**, 968–979.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms?" J. Acoust. Soc. Am. **95**, 3541–3554.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. A. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Craig, J. H., and Jeffress, L. A. (**1962**). "Effect of phase on the quality of a two-component tone," J. Acoust. Soc. Am. **34**, 1752–1760.

Dallos, P., Popper, R., and Fay, R. (**1996**). *The Cochlea* (Springer-Verlag, New York).

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Depireux, D., Simon, J., Klein, D., and Shamma, S. (**2001**). "Spectro-Temporal Response Field Characterization with Dynamic Ripples in Ferret Primary Auditory Cortex," J. Neurophysiol. **85**, 1220–1234.

Feeney, M. F. (**1997**). "Dichotic beats of mistuned consonances," J. Acoust. Soc. Am. **102**, 2333–2342.

Fleischer, V. H. (**1980**). "Subjektive grösse von unterschieden im amplituden-modulationsgrad von sinustönen (Perceived magnitude of differences in the amplitude modulation depth of sinusoids), Acustica **46**, 31–37.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1966**). "Auditory spectral filtering and monaural phase perception," J. Acoust. Soc. Am. **41**, 458–479.

Haggard, M. P., Harvey, A. D. G., and Carlyon, R. P., (**1985**). "Peripheral and central components of comodulation masking release," J. Acoust. Soc. Am. **78**, S63.

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., and Patterson, R. D. (**2002**). "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform," Speech Commun. **36**, 181–203.

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.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.

Kowalski, N., Depireux, D., and Shamma, S. (**1996**). "Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single unit responses to moving ripple spectra," J. Neurophysiol. **76**, 3503–3523.

Lyon, R., and Shamma, S. A. (**1996**). "Auditory representations of timbre and pitch," in *Auditory Computation, Springer Handbook of Auditory Research* (Springer-Verlag, New York).

Mathes, R.C., and Miller, R. L. (**1947**). "Phase effects on monaural perception," J. Acoust. Soc. Am. **19**, 780–797.

McFadden, D. (**1986**). "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," J. Acoust. Soc. Am. **80**, 1658–1667.

Meddis, R., and Hewitt, M. (**1991a**). "Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: Phase sensitivity," J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and Hewitt, M. (**1991b**). "Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J., and Schonveldt, G. P. (**1990**). ''Comodulation masking release as a function of bandwidth and time delay between one-frequency and flanking-band maskers,'' J. Acoust. Soc. Am. **88**, 725–731.
Patterson, J. H., and Green, D. M. (**1970**). ''Discrimination of transient signals having identical energy spectra,'' J. Acoust. Soc. Am. **48**, 894–905.
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.
Plomp, R. (**1967**). ''Beats of mistuned consonances,'' J. Acoust. Soc. Am. **42**, 462–474.
Rosen, S., and Baker, R. J. (**1994**). ''Characterising auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.
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.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Shamma, S. (**1985**). ''Speech processing in the auditory system: II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve,'' J. Acoust. Soc. Am. **78**, 1622–1632.
Shamma, S., Versnel, H., and Kowalski, N. (**1995**). ''Ripple Analysis in the Ferret Primary Auditory Cortex. I. Response Characteristics of Single-Units to Sinusoidally Rippled Spectra,'' J. Aud. Neurosci. **1**, 233–254.
Slaney, M., and Lyon, R. F. (**1990**). ''A perceptual pitch detector,'' Proc. Int. Conf. Acoustics, Speech, and Signal Processing 357–360.
Strickland, E., Viemeister, N. F., Fantini, D. A., and Garrison, M. A. (**1989**). ''Within- versus cross-channel mechanisms in detection of envelope phase disparity,'' J. Acoust. Soc. Am. **86**, 2160–2166.
Wang, K., and Shamma, S. A. (**1994**). ''Self-normalization and noise-robustness in early auditory representations,'' IEEE Trans. Speech Audio Process. **2**, 421–435.
Wang, K., and Shamma, S. A. (**1995**). ''Spectral shape analysis in the central auditory system,'' IEEE Trans. Speech Audio Process. **3**, 382–395.
Yang, X., Wang, K., and Shamma, S. A. (**1992**). ''Auditory representations of acoustic signals,'' IEEE Trans. Inf. Theory **38**, 824–839.
Yost, W. A., and Sheft, S. (**1989**). ''Across-critical-band processing of amplitude-modulated tones,'' J. Acoust. Soc. Am. **85**, 848–857.

Asymmetry of masking between complex tones and noise: Partial loudness

Hedwig Gockel[a)]
CNBH, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, England

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Roy D. Patterson
CNBH, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, England

(Received 29 July 2002; revised 27 January 2003; accepted 25 April 2003)

This experiment examined the partial masking of periodic complex tones by a background of noise, and vice versa. The tones had a fundamental frequency (*F*0) of 62.5 or 250 Hz, and components were added in either cosine phase (CPH) or random phase (RPH). The tones and the noise were bandpass filtered into the same frequency region, from the tenth harmonic up to 5 kHz. The target alone was alternated with the target and the background; for the mixture, the background and target were either gated together, or the background was turned on 400 ms before, and off 200 ms after, the target. Subjects had to adjust the level of either the target alone or the target in the background so as to match the loudness of the target in the two intervals. The overall level of the background was 50 dB SPL, and loudness matches were obtained for several fixed levels of the target alone or in the background. The resulting loudness-matching functions showed clear asymmetry of partial masking. For a given target-to-background ratio, the partial loudness of a complex tone in a noise background was lower than the partial loudness of a noise in a complex tone background. Expressed as the target-to-background ratio required to achieve a given loudness, the asymmetry typically amounted to 12–16 dB. When the *F*0 of the complex tone was 62.5 Hz, the asymmetry of partial masking was greater for CPH than for RPH. When the *F*0 was 250 Hz, the asymmetry was greater for RPH than for CPH. Masked thresholds showed the same pattern as for partial masking for both F0's. Onset asynchrony had some effect on the loudness matching data when the target was just above its masked threshold, but did not significantly affect the level at which the target in the background reached its unmasked loudness. The results are interpreted in terms of the temporal structure of the stimuli. [DOI: 10.1121/1.1582447]

PACS numbers: 43.66.Cb, 43.66.Dc, 43.66.Nm [MRL]

I. INTRODUCTION

Recent detection experiments have demonstrated that there are large asymmetries in the masking produced by broadband Gaussian noise, on the one hand, and tonal sounds with the same excitation patterns, on the other hand. Specifically, Gockel *et al.* (2002a) have demonstrated that noise masks harmonic complex tones much more effectively than the reverse and Krumbholz *et al.* (2001) have shown that noise masks iterated rippled noise (IRN—a stimulus which evokes a pitch sensation) more effectively than the reverse. In these experiments, when the target sound is well above threshold, a tone and a noise are perceived separately, despite the fact that both excite the same range of frequency channels, and each of the percepts seems to have its own loudness. In the present paper we report an experiment to assess whether and how the asymmetry of masking observed for masked thresholds translates to supra-threshold levels; the partial loudness of harmonic complex tones in noise, and vice versa, was determined for targets and maskers with identical bandwidth.

Partial loudness or masked loudness refers to the loudness of a target sound when presented together with a background or partial masking sound (Scharf, 1964; Zwicker and Fastl, 1999). The partial loudness of a target sound can be determined either directly by the use of magnitude estimation or magnitude production procedures, or indirectly using loudness-matching procedures (Zwicker, 1958, 1963). In a typical loudness-matching procedure, the target is presented alone in one interval and together with the masker in the other interval; its level is fixed in one interval and variable in the other. The listener adjusts the level in order to achieve equal loudness of the target in the two intervals (Zwicker, 1963; Hellman, 1972). If this is repeated for various fixed levels of the target, a masked loudness-matching function (MLMF) results, showing the level of the target alone as a function of the level of the target in the mixture at the point

[a)]Present address: MRC-Cognition and Brain Sciences Unit, 15 Chaucer Rd., Cambridge CB2 2EF, England. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

of equal loudness. The general shape of such an MLMF can be described as follows: When the target sound is close to its threshold in the masker, the level of the target in the masker is set much higher than the level of the target alone. With increasing level of the target in the masker, the matched level of the target alone also increases, but at a faster rate. At a sufficiently high level, the level of the target in the background equals that of the target alone, and this equality then persists at still higher levels (Zwicker and Fastl, 1999).

Most studies of partial loudness have investigated the MLMF for a pure tone as a function of the frequency of the tone and the bandwidth of the masking noise (Zwicker, 1963; Hellman and Zwislocki, 1964; Scharf, 1964; Stevens and Guirao, 1967; Hellman, 1970). Only a few studies have investigated the partial loudness of noise when masked by a pure tone (Zwicker, 1963; Hellman, 1972; Schroeder *et al.*, 1979). Zwicker (1963) measured the level of a narrow-band noise target required to give the same loudness as that target presented in a pure tone background, as a function of the level of the background; the frequency of the background equaled the center frequency of the noise. However, he did not measure MLMFs. Schroeder (1979) measured MLMFs for a critical-band wide noise centered at 1 kHz, masked by an 80 dB SPL sinusoid with various frequencies. The effect of the masking tone was greatest when it was centered on the noise band. Schroeder estimated the slope of the MLMF for a tone frequency of 1 kHz to be 3 dB/dB (level of noise alone plotted against level of noise presented with masker) when the level of the noise was below 80 dB SPL.

Hellman (1972) provided the sole study in which the growth of the masked loudness of a noise in a tone was directly compared with that of a tone in a noise. Note, however, that the data for the latter condition were obtained from different subjects, in an earlier study (Hellman, 1970). Hellman (1970, 1972) used a 1000-Hz pure tone and noise bands of various widths. In the 1970 paper, she investigated MLMFs for a pure tone masked by noises of various bandwidths and levels. For nearly comparable overall masker SPLs, an octave-band noise produced a steeper MLMF than a wide-band noise, and the MLMF was shifted to the right for the former. Increasing the level of a fixed-bandwidth noise masker also produced steeper, right-shifted MLMFs. Importantly, two noise maskers with different bandwidths but with levels chosen to give equal masked thresholds for the tone, produced very similar MLMFs. In the 1972 paper, Hellman reported MLMFs for noise targets of various bandwidths and various levels of the pure tone masker. The slope of the MLMF for the noise target masked by a pure tone with a frequency centered in the noise band became steeper as the bandwidth of the target noise decreased. Comparison of MLMFs for a critical-band wide noise target from the later paper, with MLMFs for the pure tone target and a critical-band wide noise masker from the earlier paper, showed that the two functions agreed reasonably well in position and slope if the level of the tone masker was about 18 dB higher than that of the noise masker, i.e., the noise was the more effective masker. Also, the masked threshold for the tone in noise was about 20 dB higher than the masked threshold for the noise in the tone. Hellman (1972) concluded that the partial loudness functions for the noise target and the tone target were ''surprisingly similar,'' if the level of the tone masker was increased relative to that of the noise masker by the amount necessary to obtain equal masked thresholds of the noise in the tone and the tone in the noise.

The present study measured MLMFs using broadband sounds as targets and maskers; the targets and maskers were complex tones and noises with identical bandwidth and near-identical excitation patterns (see below). There were four main objectives:

(1) To measure the growth of partial loudness for complex sounds more like those that are common in speech and music.
(2) To assess whether the asymmetry between the MLMFs of tones masked by noises and of noises masked by tones would be observed for complex tones and noise of equal bandwidth, or whether the previously observed asymmetry in partial loudness for tones and noise might have been dependent on some difference in the excitation patterns of the stimuli.
(3) To assess the influence of onset and offset asynchrony of the target and masker on the resulting MLMFs. In a variety of paradigms, onset and offset asynchrony have been shown to be powerful cues for perceptual segregation of two concurrent sounds (see, e.g., Darwin and Carlyon, 1995). Therefore, one might expect the target sound to be heard out more clearly and possibly to be louder if the target and masker start asynchronously than if they start synchronously. In past experiments on partial loudness, asynchronous onsets have generally been used in order to facilitate loudness comparisons between the target in the background and the target alone. However, we are not aware of any study quantifying the effect of onset asynchrony on partial loudness.
(4) To investigate the influence of the starting phases of the components of the complex tones, which affects the degree of envelope modulation, on the MLMFs. For component phases giving a highly modulated waveform (cosine phase) one might expect a clearer and stronger pitch percept than for phases giving a less modulated waveform (random phase) (Moore, 1977; Lundeen and Small, 1984). A target with a clearer pitch, and therefore less perceptual similarity with the background noise, might be perceptually segregated from the noise more effectively than a target with a weak pitch. As in the case of onset asynchrony, a higher degree of perceptual segregation between complex tone and noise might cause the target sound to be heard as louder. The degree of envelope modulation in the internal representation of a complex tone, i.e., after basilar-membrane (BM) filtering, is dependent on the number of components interacting at the output of a given auditory filter; the more components the higher the degree of envelope modulation. Hence, for a fixed center frequency, the lower the fundamental frequency ($F0$), the more modulated is the envelope. To further check the effect of degree of envelope

modulation and of duration of the repetition period on the MLMFs, two *F*0s spaced two octaves apart were employed.

The stimuli were identical to those used by Gockel *et al.* (2002a). The *F*0 of the complex tones was either 62.5 or 250 Hz, and the components of the complex tones started in either random or cosine phase. Gockel *et al.* (2002a) found that the asymmetry of masking as measured by masked thresholds depended on the starting phases of the components, the *F*0, and, for the low *F*0, on the level of the masker. We were also interested in determining whether the same pattern of dependence would be observed for partial loudness.

II. METHOD

A. Stimuli

The sound to be matched in loudness is called the target, throughout the study. We used either a harmonic complex-tone target masked by a noise, or a noise target masked by a harmonic complex tone. The complex tones had an *F*0 of 62.5 or 250 Hz. The components were added in either random phase (RPH) or cosine phase (CPH). All tones were bandpass filtered into a frequency region from the tenth harmonic to 5000 Hz at the 3-dB down points. The slope of the filter outside the passband was 100 dB/oct. Thus, the passband was from 625 to 5000 Hz for the 62.5-Hz *F*0, and from 2500 to 5000 Hz for the 250-Hz F0. As a result, the components of the harmonic complexes were unresolved by the peripheral auditory system (Plomp, 1964; Moore and Ohgushi, 1993). The CPH complex would be expected to produce a waveform whose envelope, at the output of each auditory filter, had a higher degree of modulation than that of the RPH complex. The noise presented with a given complex tone (either as target or masker) was a Gaussian noise filtered in the same way as the given complex. The root-mean-square (rms) level of the masker was 50 dB SPL.

For each *F*0, the excitation patterns for the tone and noise stimuli were essentially identical, when calculated using the procedure described by Glasberg and Moore (1990); see Fig. 1 of Gockel *et al.* (2002a). Thus, based on the long-term excitation patterns, no partial masking asymmetry would be expected between tone targets with noise maskers and noise targets with tone maskers.

The duration of the target was always 700 ms (including ramps). The masker was gated either synchronously with the target (the ''synchronous'' condition), or its duration was 1300 ms, in which case it was gated on 400 ms before and off 200 ms after the target (the ''asynchronous'' condition). In the synchronous condition, all signals were gated with 40-ms, raised-cosine onset and offset ramps. In the asynchronous condition, the target was gated with 40-ms ramps and the masker was gated with 80-ms ramps.

The stimuli were generated digitally in advance using a sampling rate of 25 kHz. The tones were generated by adding up sinusoids with frequencies ranging from *F*0 up to 10 kHz, while the noise was generated in the temporal domain by sampling from a Gaussian distribution. Bandpass filtering was then performed with a 900-tap digital FIR filter with a linear phase response. Ten different realizations were produced for each RPH complex tone and for the Gaussian noise; one of the ten was picked at random for each presentation. The target and masker were played out through separate channels of a Tucker-Davis Technologies (TDT) DD1 16-bit digital-to-analog converter, and then lowpass filtered at 10 kHz (TDT FT6-2). They were separately attenuated using two programmable attenuators (TDT PA4) and then added (TDT SM3). Stimuli were fed to a headphone buffer (TDT HB6) and presented monaurally via headphones with a diffuse-field response (AKG K 240 DF). Subjects were seated individually in an IAC double-walled sound-attenuating booth.

B. Procedure

A two-alternative forced-choice task was used. The target alone and the target plus masker were presented monaurally, in regular alternation, with 200-ms silent intervals between successive sounds. The alternating intervals were each marked by a light, and subjects were required to indicate in which of the two intervals they heard the target as louder by pressing the button underneath the corresponding light on a response box. Within a given run, either the target alone or the target within the mixture was fixed in level, and the level of the target in the other interval was varied to determine the level corresponding to equal loudness of the target in the two intervals. The range of levels used for the fixed target and the range of starting levels for the variable target were determined after a preliminary experiment. The levels of the fixed target in the masker were chosen so as to cover the range from ''just above masked threshold'' to ''sounds as loud as target presented alone.'' The levels of the fixed target alone were chosen so that the highest level was the same as that used for the target in the masker, and the lowest level corresponded to the level matching in loudness the lowest level of the target in the masker. On average, ten fixed target levels were used in each condition. The starting level of each variable target was chosen randomly from a range of levels. This range of levels was different for each type of target and each fixed target level; the range was symmetrical about the matched level obtained in the preliminary experiment for the given fixed target level.

When the fixed target was judged to be the louder of the two, the target with the variable level was increased in level. When the variable target was judged to be louder, it was decreased in level. The sound level was changed only between stimuli. When no button was pressed, sound presentation continued without any change. A change in the button pressed defined a turnpoint. Initially, the target level was increased or decreased in 5-dB steps. After two turnpoints the step size was reduced to 3 dB, and after four turnpoints to 1 dB. Subjects were encouraged to bracket the point of equal loudness several times, by going from ''target alone is louder'' to ''target within the mixture is louder,'' before making the fine adjustment to achieve equal loudness. When they were satisfied with the match, they pressed a third (green) button, which stopped the run. This ''satisfied button'' press was only accepted after the final step size was reached. If

subjects pressed it earlier, it was ignored, and sound presentation continued without a change. The level of the variable sound at the point when the run was stopped was taken as the matching level.

Subjects were asked only to press the ''satisfied button'' if they were able to match the loudness of the target alone and the target heard in the mixture, i.e., when they actually heard a sound in the mixture which resembled the target when heard alone. Given this instruction, subjects did not produce matches in some cases, e.g., when the noise target was presented at a very low fixed level with the tone masker. In such a case, the mixture was perceived as a rough tone, i.e., one sound object. The roughness was perceived as an attribute of the tone and, by itself, had no loudness which could be judged; only the tone had a loudness. Similarly, if the noise target alone was presented at a low fixed level, the subjects sometimes could not adjust the level of the noise target in the tone masker in such a way that they actually heard the noise as a sound object in the mixture with the same soft loudness as the noise alone. By the time the variable level of the noise target in the mixture was high enough to enable the subject to hear the noise as such, it was louder than the noise alone. If a match in loudness was deemed to be impossible, subjects pressed a fourth (red) button. This ''impossible button'' press also was only accepted after the final step size was reached. If subjects pressed it earlier, it was ignored, and sound presentation continued without a change.

The total duration of a single session was about 2 h, including rest times. The conditions were presented in a counterbalanced order. Half of the subjects started with the harmonic complex tone as target, and half with the noise as target. In each of these conditions, for half of the subjects the complex tone was presented first in RPH and then in CPH, and for the other half the order was reversed. The two onset conditions were alternated, as was the choice of the target with the fixed level (target alone, or target with the masker). The level of the fixed target was selected randomly among the ten or so possible levels (one threshold was obtained for each condition in turn, before additional measurements were obtained in any other condition). For each subject and condition four measurements were obtained. This took about 13 sessions.

After the main experiment was completed, absolute threshold was measured for each signal, and masked threshold was measured for each target in each background, using a two-interval two-alternative forced-choice task and a three-down one-up adaptive procedure. Four threshold estimates were obtained for each condition and each subject; the mean of the four is reported. Masked thresholds were obtained only in the condition with asynchronous onsets, as we showed previously that masked thresholds for these stimuli were not affected by onset asynchrony (Gockel *et al.*, 2002a).

C. Subjects

Eight subjects participated overall. Half of them were tested using the low $F0$ and half using the higher $F0$. Their ages ranged from 20 to 22 years, and their quiet thresholds at audiometric frequencies between 500 and 5000 Hz were better than 15 dB HL. Subjects had 8 h of practice before the experiment proper was started. No feedback was given following each trial, but during practice subjects were informed about their ''performance,'' i.e., the degree to which their MLMFs were monotonic.

III. RESULTS

In what follows, we use the following six-letter system of abbreviations. The first letter (t or n) indicates whether the target was a complex tone or a noise. The second, third, and fourth letters indicate whether the tone had components in cosine phase (CPH) or random phase (RPH), regardless of whether the tone was the target or masker. The fifth and sixth letters indicate the gating condition: synchronous onsets (so) or asynchronous onsets (ao). For example, tRPHso indicates the tone target in RPH with synchronous onsets of target and masker, while nCPHao indicates the noise target with the CPH masker and asynchronous onsets of target and masker.

Initially, the data were analyzed individually for each subject. Figure 1 shows an example of individual data for the 62.5-Hz $F0$.[1] Panels (a)–(d) show conditions with a tone target and panels (e)–(h) show conditions with a noise target. Each symbol represents one loudness match between the target alone (level on the *y* axis) and the target in the masker (level on the *x* axis). Open triangles indicate matches where the level of the target alone was varied; these matches vary along the *y* axis. Open circles indicate matches where the level of the target in the masker was changed; these matches vary along the *x* axis. The asterisk shows the value of masked threshold (abscissa) and threshold in quiet (ordinate) for the target. Note that there were some trials in which the level of the fixed target was lower than any level plotted in the graph; trials where the subject deemed a match in loudness to be impossible are not plotted. Also, if, for a given fixed level, a match was obtained in one trial and was deemed to be impossible in the remaining three trials for that condition, then the single ''successful'' match was discarded. The dotted diagonal line indicates the expected matching values if the loudness of the target in the masker were unaffected by the masker.

To derive a simple description of the MLMFs, the data were fitted with two straight lines. One line was the diagonal with slope$=1$ (the dotted lines in the figures). The other was a line with slope>1, described by two parameters, its slope and its point of intersection with the diagonal. Each data point was ''assigned'' either to the variable-slope line or to the diagonal. The fitting procedure worked in the following way. Starting values were chosen for the slope and point of intersection of the variable line. Matches from trials where the fixed level of the target was higher than the intersection point were assigned to the diagonal; remaining matches were assigned to the variable line. The slope and intersection were then adjusted (the assignment of points to the two lines changed correspondingly) so as to minimize the sum of the squared distances of the matches from the variable line and from the diagonal. For matches from trials where the level of the target alone was varied, the vertical distances were minimized. For matches where the level of the target in the mix-

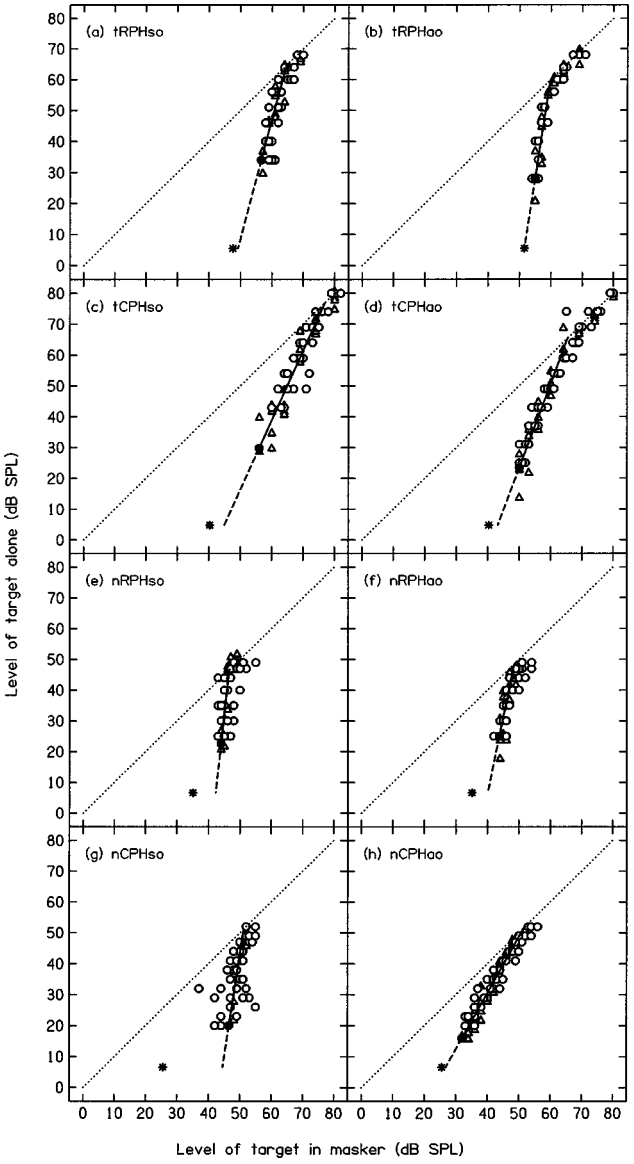


FIG. 1. Results of one subject for $F0 = 62.5$ Hz. Each panel shows results for one condition, as indicated in the key. The first letter of the key (t or n) indicates whether the target was a complex tone or a noise. The second, third, and fourth letters indicate whether the tone had components in cosine phase (CPH) or random phase (RPH), regardless of whether the tone was the target or masker. The fifth and sixth letters indicate synchronous onsets (so) or asynchronous onsets (ao). Each symbol represents one loudness match between the target alone and the target in the masker. Open triangles indicate matches from trials where the level of the target in the masker was fixed. Open circles indicate matches from trials where the level of the target alone was fixed. Note that the number of matches is greater than the number of symbols shown, because some symbols are obscured by others. Also, the number of triangles is greater than appears, since many triangles are covered by circles. The asterisk indicates the masked threshold (abscissa) and the threshold in quiet (ordinate) of the target. See text for an explanation of the fitted solid line, and its extension (dashed line) down to the absolute threshold.

ture was varied, the horizontal distances were minimized. The two types of minimization were performed simultaneously. The results for the example subject chosen are shown in Fig. 1 as solid lines.

The intersection of each fitted line with the diagonal represents the target level at which the masker ceased to have a partial masking effect. The filled circle on the lower end of each line represents the lowest level at which the subject was able to perform the task (on at least two out of four trials) for that condition. To determine this level, we considered the lowest level which was matched both when the target alone was fixed and when the target in the masker was fixed. The projection of these two matches onto the fitted line (in the *y* direction for the former and the *x* direction for the latter) was determined, and whichever match gave an intersection at the lower point on the line was taken as the lowest matching level.

The dashed line which continues from the fitted line below the lowest matched level projects to a *y* value which is identical to the measured threshold in quiet of the target. The corresponding *x* value gives the predicted masked threshold for the target under two assumptions: (1) At masked threshold the target has the same loudness as at absolute threshold (Moore *et al.*, 1997). (2) If it had been possible to measure MLMFs at lower levels, they would have continued as straight lines. The validity of these assumptions is discussed in Sec. III E.

To enable a representative comparison of the MLMFs across conditions, an average fitted line was calculated from the four lines fitted to the individual subject data in a given condition. This average fitted line represents the MLMF of a given condition for all subjects, and it was derived in the following way. First, its slope was set equal to the geometric mean of the slopes of the individually fitted lines. Second, its ''center of gravity'' was set to the point (Xg, Yg) corresponding to the mean of the ''centers of gravity'' of the individually fitted lines, where the center of gravity of an individually fitted line (Xi, Yi) is defined in the following way: Xi corresponds to the mean of the minimum *x* value on the fitted line (i.e., the lowest matched *x* level) and the *x* value at its intersection with the diagonal, and Yi corresponds to the mean of the minimum *y* value on the fitted line (i.e., the lowest matched *y* level) and the *y* value at its intersection with the diagonal. Finally, the length of the ''average'' fitted line was determined as two times the (vector) length from its center of gravity (Xg, Yg) to its intersection with the diagonal. Figure 2 shows the resulting averaged MLMFs for the 62.5-Hz *F*0 (a, top) and the 250-Hz *F*0 (b, bottom). The symbols close to the *x* axis give absolute thresholds (*y* axis) and masked thresholds (*x* axis) for the corresponding conditions, averaged across subjects.

A. Asymmetry of masking

For both *F*0s, an asymmetry of partial masking is clearly apparent; all of the fitted lines for the tone targets (open symbols) lie to the right of the fitted lines for the noise targets (solid symbols). This indicates that the asymmetry of partial masking reported by Hellman (1972) for pure tones and noise also holds for complex tones and noise with identical excitation patterns.

To quantify these ''offsets'' we used the *x*-axis values at the ''center of gravity,'' Xg, of the fitted lines averaged across all subjects. The value of this parameter is shown for each condition in Fig. 3. There was no consistent effect of onset asynchrony on Xg, so the results were averaged for synchronous and asynchronous onsets. For the 62.5-Hz *F*0,

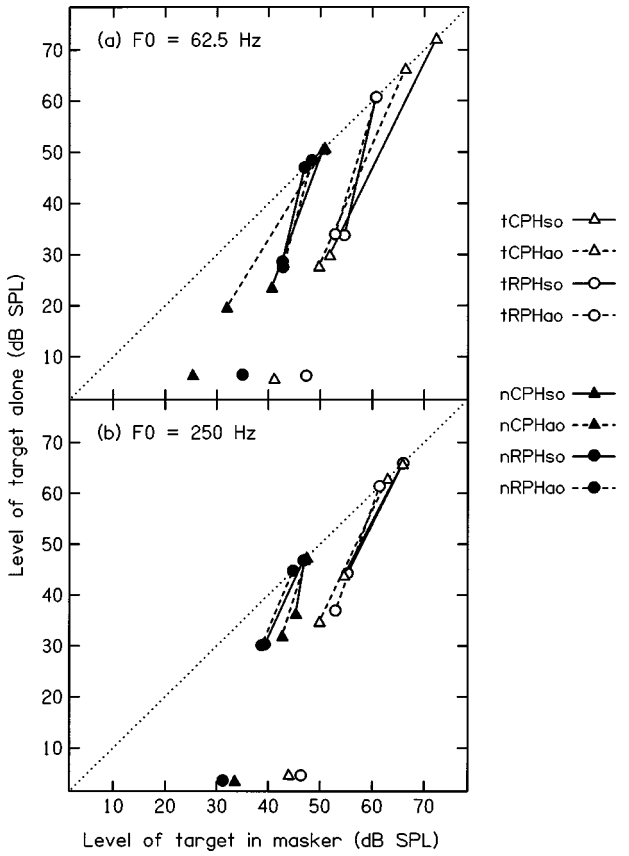


FIG. 2. Panel (a) shows the fitted lines representing the loudness-matching functions obtained for the 62.5-Hz *F*0, averaged across subjects. See the text for an explanation of how the averaged fitted lines are derived from the individual fits. The open symbols at the endpoints of the lines indicate conditions where the target was a complex tone. The solid symbols at the endpoints of the lines indicate conditions where the noise was the target. Triangles and circles indicate conditions where the components of the complex tone had cosine and random starting phases, respectively. Solid lines and dashed lines indicate synchronous and asynchronous onsets/offsets of target and masker, respectively. The individual symbols close to the *x* axis show absolute thresholds (*y* axis) and masked thresholds (*x* axis) for the corresponding conditions, averaged across subjects for the asynchronous condition only. Panel (b) shows corresponding results for $F0 = 250$ Hz.

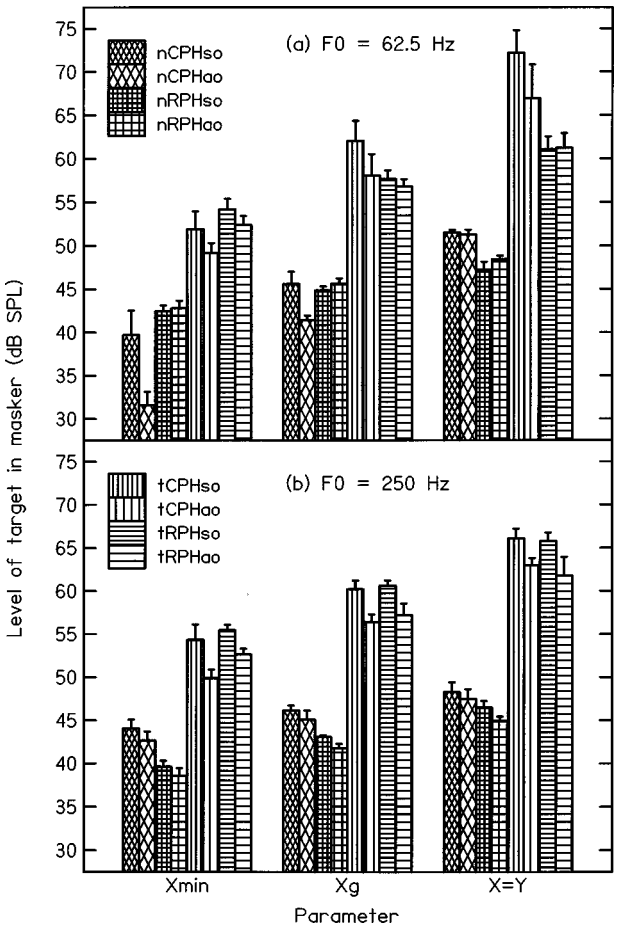


FIG. 3. Mean values of parameters describing the lines fitted to the individual data: (1) Xmin, which is the lowest level of the partially masked sound for which each subject was able to perform the task; (2) Xg, which is the *x*-axis value of the center of gravity of the average fitted line; and (3) the intersection point with the diagonal, denoted here X=Y, which gives an estimate of the level of the partially masked sound at which it reaches its unmasked loudness. The upper and lower panels show values for the 62.5- and 250-Hz *F*0s, respectively. The condition is indicated by the keys in the upper and lower panels. Error bars indicate one standard error (SE).

the effect of type of target (tone versus noise) was 16.5 dB for CPH (mean of Xg for conditions tCPHso and tCPHao minus mean of Xg for conditions nCPHso and nCPHao) and 12.0 dB for RPH (mean of Xg for conditions tRPHso and tRPHao minus mean of Xg for conditions nRPHso and nRPHao), i.e., the noise was a more efficient masker than the tones and this asymmetry of partial masking was greater for CPH than for RPH. These asymmetries were very similar to the corresponding asymmetries for the masked thresholds, shown at the bottom of each panel in Fig. 2. The latter were 15.8 (threshold of CPH tone masked by noise minus threshold of noise masked by CPH tone) and 12.4 dB (threshold of RPH tone masked by noise minus threshold of noise masked by RPH tone), respectively. For the 250-Hz *F*0, the effect of type of target was 12.7 dB for CPH and 16.5 dB for RPH, i.e., the asymmetry of partial masking was greater for RPH than for CPH, the opposite of what was found for $F0 = 62.5$ Hz. Again, the asymmetries were similar to the corresponding asymmetries for the masked thresholds, which were 10.5 and 15.2 dB, respectively. The correspondence between the partial masking effects, quantified as the horizontal shift of the center of gravity, and the masked thresholds is consistent with the findings of Hellman (1972) for the asymmetry of masking between sinusoidal tones and bands of noise, as described in the Introduction.

To assess the significance of differences between conditions, several within-subject analyses of variance (ANOVAs) were performed, based on various features of the lines fitted to the data for the individual subjects. For brevity, in what follows, only significant effects ($p<0.05$) will be discussed. To investigate the asymmetry of masking, separate ANOVAs were performed for CPH and RPH, using the factors type of target (tone or noise) and onset asynchrony. The intersection with the diagonal indicates the level of the target in the background required for the background no longer to affect loudness. The mean value of this parameter for each condition is shown in Fig. 3 by X=Y. There were clear effects of type of target on the intersection point, for both *F*0s; the *x* and *y* values at the intersection points were larger for the tone targets than for the noise targets. ANOVAs based on the intersection points showed a significant effect of whether the

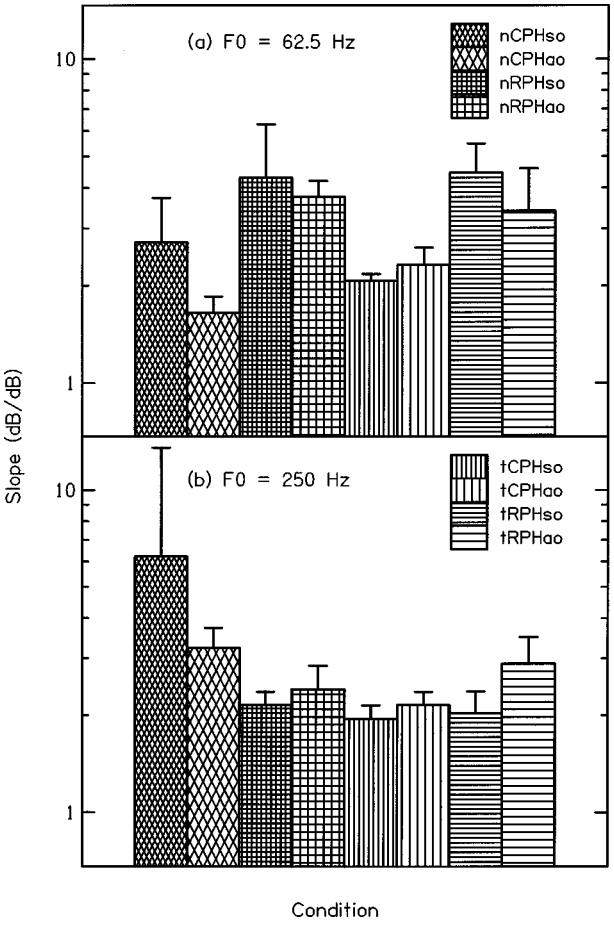


FIG. 4. Geometric mean slopes (and corresponding SEs) of the lines fitted to the individual data. The upper and lower panels show values for the 62.5- and 250-Hz *F*0s, respectively.

RPH tone was the target or masker; $F(1,3)=767.7$, $p<0.001$ at $F0=62.5$ Hz and $F(1,3)=259.4$, $p<0.001$ at $F0=250$ Hz. ANOVAs also showed a significant effect of whether the CPH tone was the target or masker; $F(1,3)=43.2$, $p=0.007$ at $F0=62.5$ Hz and $F(1,3)=256.0$, $p<0.001$ at $F0=250$ Hz. The lowest points on the fitted lines in Fig. 2 indicate the lowest matched levels. The mean *x*-axis value at the lowest points is shown as Xmin in Fig. 3. The large asymmetry of masking was also reflected in those; ANOVAs based on the Xmin values gave highly significant effects of type of target, Xmin being larger for tone targets than for the noise targets.

B. Phase effects

To investigate the effects of phase, separate ANOVAs were conducted for the tones as targets (with factors target phase and onset asynchrony) and for the noise as target (with factors background phase and onset asynchrony). ANOVAs based on (the logarithms of) the slopes of the fitted lines revealed no significant effects for $F0=250$ Hz, but several significant effects for $F0=62.5$ Hz. The geometric mean slopes of the fitted lines for each condition are shown in Fig. 4. For the tone targets only, there was a significant effect of phase; $F(1,3)=36.7$, $p=0.009$. The geometric mean slope for RPH (3.89) was greater than for CPH (2.19). An ANOVA based on the values of the intersection points (X=Y in Fig. 3) for the tone targets with $F0=62.5$ Hz also revealed a significant effect of phase; $F(1,3)=13.2$, $p=0.036$. The intersection point was lower for RPH (61.2) than for CPH (69.6). Taken together, these findings indicate that the loudness of the tone target in the noise background grew more rapidly with increasing target level for the RPH than for the CPH target.

An ANOVA based on (the logarithms of) the slopes for the noise targets alone (for $F0=62.5$ Hz) gave a significant effect of background phase; $F(1,3)=88.7$, $p=0.003$. The geometric mean slope for the RPH background (4.0) was greater than for the CPH background (2.11). For relatively low levels of the noise target in the masker, the matching level of the target alone was lower for the RPH background, while for higher levels, the matching level of the target alone was lower for the CPH background. An ANOVA based on the values of the intersection points for the noise targets (also for $F0=62.5$ Hz) revealed a significant effect of background phase; $F(1,3)=52.0$, $p=0.005$. The intersection point was lower for the RPH background (47.9) than for the CPH background (51.4). For low levels of the noise target in the RPH background, subjects reported that the target was hard to hear as a separate sound from the background; consistent with this, an ANOVA based on the Xmin values for the noise target alone (also for $F0=62.5$ Hz) showed a significant effect of background phase; $F(1,3)=10.8$, $p=0.046$. The Xmin value was larger for RPH (42.6 dB) than for CPH (35.6 dB).

For $F0=250$ Hz and for noise targets only, an ANOVA based on the Xmin values gave a significant effect of phase of the background; $F(1,3)=79.2$, $p=0.003$. The mean Xmin value was 43.4 dB for CPH and 39.1 for RPH (see Fig. 3). This asymmetry is in the opposite direction to that found for $F0=62.5$ Hz, but the asymmetries at both frequencies are consistent with those observed for the masked thresholds of noise targets in CPH or RPH maskers.

C. Onset effects

ANOVAs based on the Xmin values gave several significant effects related to the type of onset. For $F0=62.5$ Hz, considering the data for CPH only (as either target or masker), there was a significant effect of onset asynchrony [$F(1,3)=10.35$,$p=0.049$]; the mean Xmin value was lower for the asynchronous onsets (40.4 dB) than for the synchronous onsets (45.8 dB). There was also a significant interaction of type of target and onset asynchrony; $F(1,3)=17.55$, $p=0.025$. This reflects the fact that there was a relatively large effect of asynchrony for the noise as target, but only a small effect for the CPH tone as target (see Fig. 3). For the noise target, the asynchrony seemed to help the subjects in hearing the target as a separate sound from the background, when the target-to-background ratio was relatively low. When the onsets were synchronous, the noise target tended to be heard as a roughness of the background tone rather than as a separate sound.

For $F0=250$ Hz, considering the data for RPH only (as either target or masker), an ANOVA based on the Xmin val-

TABLE I. Results of the chi-square tests used to compare, for a given pair of conditions, the number of matches above and below the single line fitted to the matches for those two conditions. The tests were run for each subject individually. The top half is for the 62.5-Hz *F*0, the lower half for the 250-Hz *F*0. The condition before the ">" had a significantly greater proportion of matches above the fitted line than the condition after the ">." The number of subjects (out of four) who showed a given significant effect ($p<0.05$) is indicated.

$F0=62.5$ Hz, comparison of synchronous with asynchronous conditions				
Effect	tRPH..	tCPH..	nRPH..	nCPH..
A0>S0	2	2	···	4
S0>A0	···	···	3	···
$F0=62.5$ Hz, comparison of RPH with CPH conditions				
Effect	t...ao	t...so	n...ao	n...so
RPH>CPH	2	3	···	3
CPH>RPH	···	···	3	···
$F0=250$ Hz, comparison of synchronous with asynchronous conditions				
Effect	tRPH..	tCPH..	nRPH..	nCPH..
A0>S0	3	3	1	1
S0>A0	···	1	1	1
$F0=250$ Hz, comparison of RPH with CPH conditions				
Effect	t...ao	t...so	n...ao	n...so
RPH>CPH	···	1	3	3
CPH>RPH	1	···	···	···

ues gave a significant effect of onset asynchrony; $F(1,3)=12.2$, $p=0.04$, asynchronous onsets leading to a lower Xmin value (45.6 dB) than synchronous onsets (47.6 dB). However, the effect was small.

D. Individual subject data

Inspection of the data for individuals showed sizable effects of onset asynchrony and component phase for some subjects but not for others. To assess the significance of effects for individual subjects, pairs of conditions were compared in the following way. The matches for a given pair of conditions were fitted as if they were all from the same population, i.e., fitted with a single sloping line (plus the diagonal), using the method described earlier. Chi-square tests were then used to compare the proportion of matches falling above and below the fitted line for each condition. On the null hypothesis that the two conditions do not differ, the proportion should be the same for the two conditions. The significant outcomes of these analyses ($p<0.05$) are shown in Table I.

Generally, the effects which occurred most consistently across subjects were also significant in at least one of the within-subject ANOVAs. For example, for the noise target in the CPH background with $F0=62.5$ Hz, at a given target-to-background ratio the matching levels of the target alone were higher for asynchronous than for synchronous onsets.

E. Discrepancy between predicted and obtained masked thresholds

Consider now the degree of correspondence between the obtained masked thresholds and the masked thresholds predicted from projection of the fitted lines. Recall that the predicted thresholds are based on two assumptions: (1) At masked threshold the target has the same loudness as at absolute threshold (Moore *et al.*, 1997). (2) If it had been possible to measure MLMFs at lower levels, they would have continued as straight lines.

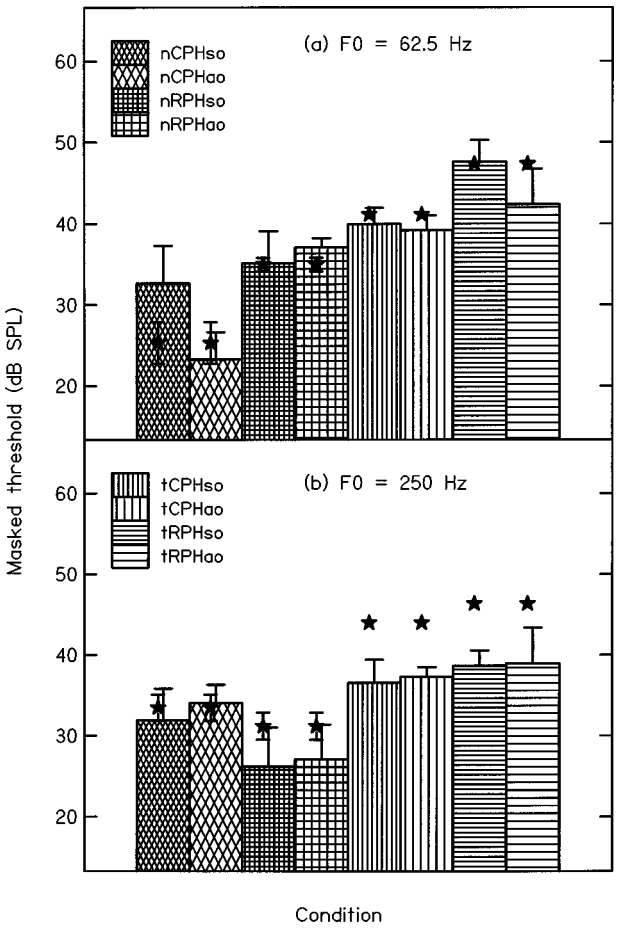


FIG. 5. The bars show mean masked thresholds predicted by extrapolation to absolute threshold of the lines fitted to the data for individual subjects, with associated SEs. The stars show the mean measured masked thresholds, with associated SEs. The upper and lower panels show values for the 62.5- and 250-Hz *F*0s, respectively.

Figure 5 compares the mean values of the predicted (bars) and obtained (stars) thresholds. For $F0=62.5$ Hz, there were no consistent differences between obtained and predicted thresholds for any condition. For $F0=250$ Hz, the predicted thresholds for the tone targets (both CPH and RPH) were consistently below the obtained thresholds, for both synchronous and asynchronous onsets. A chi-square test based on the data for all conditions with tone targets revealed a significant difference between obtained and predicted thresholds ($p<0.001$, two-tailed). For the noise target, there was no consistent difference between obtained and predicted thresholds. The effect for the tone targets is consistent with what has been observed in some previous studies of partial masking using pure tone targets and noise maskers; the MLMF becomes steeper when the tone level in the noise is very close to masked threshold (Zwicker, 1958). However, no such effect can be seen in our results, as subjects were not able to make loudness matches when the tone target was less than 6–7 dB above its masked threshold in the noise.

IV. DISCUSSION

A. Asymmetry of masking, *F*0, and phase effects

The largest effect in our results was the asymmetry of partial masking between the tone and noise stimuli; noises partially masked complex tones more effectively than complex tones masked noises. The magnitude of the asymmetry was similar to that found for masked thresholds, as was previously reported to be the case for sinusoidal targets and noise maskers (Hellman, 1972). It should be noted, however, that the horizontal separation between the MLMFs for conditions involving different targets does not generally indicate the difference in target-to-background ratio needed to achieve equal loudness. This is so because, for a fixed overall level, the CPH target, the RPH target, and the noise target typically differ in loudness (Gockel *et al.*, 2002b).

It is useful to consider whether the mechanisms that have been proposed to account for the asymmetry of masked threshold, reviewed in Gockel *et al.* (2002a), can also account for the asymmetry of partial loudness. One explanation for the asymmetry in masked threshold is based on the idea that the target is detected by comparing overall level across the intervals in a forced-choice trial. When the masker is a tone, the masker level does not fluctuate from one interval to the next. This makes it easy to detect a small increment in level produced by adding a noise target to the tone masker. However, when the masker is a noise, the level fluctuates from one interval to the next, and this makes it more difficult to detect the change in level produced by adding a tone to the noise (Bos and de Boer, 1966; Hellman, 1972). It is difficult to see how this explanation could apply to partial loudness, at least for target levels that are well above masked threshold. In order to perform the loudness-matching task, the target has to be segregated from the background, and the overall level of the stimuli is not directly relevant. It is clear that subjects were not matching the overall loudness of the target plus the background, since the matching level of the target alone was often well below that of the background.

It has been proposed that the relative bandwidth of the target and masker is important in determining the asymmetry of masked thresholds. Hall (1997) varied the bandwidth of both the target and the masker, from 0 to 256 Hz; the 0-bandwidth stimulus was a pure tone, while other stimuli were noise-like. For a given masker bandwidth, masked thresholds were highest, and essentially constant, for target bandwidths smaller than or equal to that of the masker. For target bandwidths exceeding that of the masker, thresholds decreased with increasing target bandwidth and with decreasing masker bandwidth. Hall concluded that, for target bandwidths smaller than or equal to the masker bandwidth, the results could be explained by a model based on long-term average energy, and that, for greater target bandwidths, the temporal structure of the target must be used as an additional cue. However, the previous results of Gockel *et al.* (2002a) and Krumbholz *et al.* (2001) do not support the idea that relative bandwidth is the critical variable. The present results extend this conclusion from masked thresholds to partial loudness.

Another explanation for the asymmetry of masked threshold is that, when a tone masker is used, the detection cue is the *within-interval* random fluctuation in level introduced by the noise target (Moore *et al.*, 1998); the masker alone has only regular within-interval fluctuation associated with its periodicity, so this cue is highly salient. A noise masker fluctuates randomly in level within each interval, so the level-fluctuation cue is not available in this case, except to the extent that the tone target *reduces* the amount of random fluctuation (Moore, 1975; Richards, 1992; Richards and Nekrich, 1993). This sort of information is reflected in the envelope spectra of the stimuli and is used in models for masked thresholds incorporating a modulation filter bank (Derleth and Dau, 2000; Verhey, 2002).

An explanation similar to this could account for the asymmetry of partial loudness. Presumably, what makes a noise sound noise-like are the random fluctuations in amplitude that (for a broadband noise) are independent in different frequency regions. When a noise target is presented in a tone background, it introduces audible random fluctuations even when it is at a low level relative to the background. When the fluctuations become sufficiently large, the noise becomes audible as a separate sound. Put another way, the steady nature of the background may make it relatively easy for the auditory system to ''cancel'' the effect of the background, the residue being the target noise. In this context, it is noteworthy that the intersection points for the noise targets occurred at or slightly below the level of the background (50 dB SPL).

A tone signal is characterized by its regular periodic nature, which makes it appear to be steady, rather than fluctuating. In auditory models such as the Auditory Image Model (Patterson *et al.*, 1995), periodic sounds produce stable auditory images. The regular periodicity of a tone target is easily disrupted by a background noise. Even when a tone target has a level well above that of the background noise, its periodicity is disturbed to some extent. It may be that a tone target in noise only reaches its unmasked loudness when the target-to-background ratio is so high that the noise produces a negligible disruption of the regular periodicity of the target. Consistent with this, the level of the tone target at the intersection was typically 10–20 dB above the noise background level.

For the 62.5-Hz tone targets, the intersection point was higher for CPH than for RPH. This may have happened because the noise introduced audible fluctuations in amplitude during the dips of the target waveforms evoked on the basilar membrane. The relatively long dips in the 62.5-Hz targets may have led to these fluctuations being sufficiently salient to reduce the loudness of the CPH target. Perhaps the CPH target only reached its unmasked loudness when the target-to-background ratio was high enough that the noise had a negligible effect on the dips in the target waveform evoked on the basilar membrane.

For the 250-Hz *F*0, the intersection point was similar for the CPH and RPH tone targets. This may reflect a reduced contribution of information from the dips of the CPH target, due to the limited temporal resolution of the auditory system; the 4-ms period of the 250-Hz target is shorter than the estimated duration of the ear's ''temporal window'' (Moore *et al.*, 1988; Oxenham and Moore, 1994). The difference be-

tween the two *F*0s may also reflect the fact that, for the lower *F*0, four times as many components would fall within the passband of a given auditory filter as for the higher *F*0. Thus, the envelope of the CPH tone with the lower *F*0 would exhibit greater modulation after auditory filtering than that of the CPH tone with the higher *F*0. At the higher *F*0, the degree of modulation might not have been sufficient for the dips in the CPH waveform to play a strong role.

Several other significant effects of the phase of the complex tone were found. For $F0 = 62.5$ Hz, the slopes of the MLMFs for the tone targets were steeper for RPH than for CPH. As in the case of the intersection points (lower for RPH than for CPH), this also may indicate that the minima in the basilar-membrane waveforms evoked by the stimuli need to be at a relatively high level relative to the noise for the noise to have a negligible effect on loudness. For $F0 = 62.5$ Hz, the slopes of the MLMFs for the noise targets were steeper for a RPH background than for a CPH background. In this case, there was a ''crossover'' effect. For low target-to-background ratios, the noise target was more reduced in loudness (relative to its unmasked value) in the RPH background than in the CPH background. At higher ratios, the reverse was true. The effect at low levels may reflect the perceptual similarity of the noise target and the RPH background; the RPH background at the low *F*0 sounded rather noiselike. This made it difficult to hear the noise target as a separate sound from the background. Presumably, for low target-to-background ratios, some of the neural activity evoked by the noise target was ''assigned'' to the background when loudness was judged. The reversed effect at high target-to-background ratios may be related to the relatively large peaks evoked by the CPH background on the basilar membrane. These large peaks might produce a partial masking effect even for relatively high target-to-background ratios.

B. Onset effects

The largest effect of onset asynchrony occurred for the noise target presented in the 62.5-Hz *F*0 CPH background. When the target-to-background ratio was relatively low, the loudness of the target was reduced less (relative to its unmasked value) in the asynchronous condition than in the synchronous condition. Also, the lowest level of the noise target in the CPH background at which subjects were able to perform the matching task was lower for asynchronous than for synchronous onsets. The asynchrony seemed to help the subjects in hearing the target as a separate sound from the background, when the target-to-background ratio was relatively low. When the onsets were synchronous, the noise target tended to be heard as a roughness of the background tone rather than as a separate sound. For the 250-Hz *F*0, the lowest target-to-background ratio of either the tone or noise target at which subjects could perform the task was lower for asynchronous than for synchronous presentation. However, this effect was significant only in the presence of the RPH complex tone. Again, it appears that asynchronous presentation made it easier to hear the target as a separate sound from the background.

It is noteworthy that there were some conditions where, for target levels just above those corresponding to the intersection point, the target alone was consistently matched to a level slightly greater than that of the target in the mixture. It was as if part or all of the energy in the background was assigned to the target. This happened most consistently for the noise target in the RPH background with synchronous onsets, and it occurred for both *F*0s. Onset asynchrony typically abolished this effect.

The effects of onset asynchrony may reflect perceptual grouping processes; onset asynchrony can be a powerful cue for the segregation of one sound from another (Darwin and Carlyon, 1995). Adaptation in the auditory system may influence the grouping process and may also have a direct effect on the loudness of the target. The neural response to the background presumably declines with time, and this may lead to a relatively strong onset response to the target when it is gated on after the background (Smith, 1979; Viemeister, 1980). As noted in the Introduction, most previous studies of partial masking have used asynchronous onsets of the target and background, to make it easier for subjects to ''hear out'' the target. Our results indicate that asynchronous presentation can lead to results somewhat different from those obtained with synchronous gating. However, for most of the conditions of our experiment, the effects of asynchrony were small. Asynchrony did not have a significant effect (at the 0.05 level) on the level at which the target in the background reached its unmasked loudness (the intersection).

C. Model considerations

Most models of loudness perception derive loudness from the power spectrum of the sound of interest (Fletcher and Munson, 1933; Zwicker, 1958; Zwicker and Scharf, 1965; Stevens, 1972; Moore *et al.*, 1997; Glasberg and Moore, 2002). None of these models would be able to predict the asymmetry between complex tones and noise in our partial loudness data, as the tones and noise had identical excitation patterns.

As mentioned above, the basic aspects of asymmetry of masked thresholds for pure tones in noise (and vice versa), and for noises with different bandwidths, can be explained by models based on the concept of a modulation filter bank (Derleth and Dau, 2000; Verhey, 2002). Similarly, the Auditory Image Model has been successfully used to predict the asymmetry of masked threshold for IRN in noise and vice versa (Krumbholz *et al.*, 2001). It might be possible to extend such models, in which temporal information is preserved, to account for the asymmetry in partial loudness of noise and complex tones. To date, these models have mainly been used for predicting threshold or pitch, and not for predicting loudness. However, a decision statistic based on the similarity or difference between the internal (model) representation resulting from the interval containing the target only and that resulting from the interval containing the target plus masker might be used to predict partial loudness judgements.

Gockel *et al.* (2002a) pointed out that, in order to account for the strong level dependence which was apparent in the masked threshold data for the noise target in the CPH

masker at $F0 = 62.5$ Hz, those models would need to include additional processes, such as BM compression. In the present experiment, the level of the background was fixed at 50 dB SPL. Thus, a model for the present data *only* might not require inclusion of BM compression. However, as the size and direction of the phase effects in the asymmetry of partial masking have been found to be closely linked to those observed for masked thresholds, a model to account for partial loudness data at various background levels would probably require inclusion of BM compression.

V. SUMMARY AND CONCLUSIONS

We have demonstrated a strong asymmetry of partial masking between periodic complex tones and bands of noise, filtered so that the target and background always fell into the same frequency region; for a given target-to-background ratio, the partial loudness of a complex tone in a noise background was reduced more (relative to its unmasked value) than the partial loudness of a noise in a complex tone background. Expressed as the difference between the *x*-axis value of the center of gravity of the fitted lines, the asymmetry typically amounted to 12–16 dB. These findings extend previous research on the asymmetry of partial masking between pure tones and narrow bands of noise. Our complex tones contained only high, unresolved harmonics, and excitation patterns calculated from the power spectra of the stimuli were almost identical for the targets and the backgrounds. Thus, the results cannot be explained by models of partial masking based on the calculation of excitation patterns from the long-term or short-term power spectra (Zwicker and Scharf, 1965; Moore *et al.*, 1997; Zwicker and Fastl, 1999; Glasberg and Moore, 2002). The study clearly shows that the asymmetry arises from differences in the temporal structure of the complex tone and noise stimuli.

When the *F*0 of the complex tone was 62.5 Hz, the asymmetry of partial loudness was greater for CPH than for RPH. The same effect was found for the masked thresholds. The phase effect presumably reflects the fact that the waveforms evoked on the basilar membrane have a higher peak factor for CPH than for RPH. When the *F*0 of the complex tone was 250 Hz, the asymmetry of partial loudness was opposite to that found for $F0 = 62.5$ Hz; the asymmetry was greater for RPH than for CPH. Again, the masked thresholds showed the same pattern as for the partial loudness. The difference between the two *F*0s probably reflects the shorter period associated with the higher *F*0, which, owing to the limited temporal resolution of the auditory system, decreases the importance of information in the dips of the waveforms evoked on the basilar membrane. The difference between the two *F*0s may also reflect the fact that, for the lower *F*0, four times as many components fell within the passband of a given auditory filter as for the higher *F*0. Thus, the envelope of the CPH tone with the lower *F*0 would exhibit greater modulation after auditory filtering than that of the CPH tone with the higher *F*0.

Onset asynchrony had some effect on the loudness matching data when the target was just above its masked threshold. For some conditions, the lowest target-to-background ratio at which the task could be performed was lower for asynchronous than for synchronous onsets; the asynchrony seemed to make it easier to hear the target as a separate sound. However, asynchrony did not have a significant effect on the intersection points, which correspond to the lowest level at which the target in the background reached its unmasked loudness.

ACKNOWLEDGMENTS

This work was supported by the MRC (Grant No. G9900362). We thank Hazel Longman, Robert Loxley, Clare Palmer, and Carlen Reed-Poysden for data collection. We thank Brian Glasberg for considerable assistance in fitting the data, and Jesko Verhey for helpful comments on a previous version of the manuscript. We also thank Marjorie Leek and two anonymous reviewers for helpful comments.

[1]The complete set of individual data can be obtained from http://www.mrc-cbu.cam.ac.uk/personal/hedwig.gockel/

Bos, C. E., and de Boer, E. (**1966**). ''Masking and discrimination,'' J. Acoust. Soc. Am. **39**, 708–715.
Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
Derleth, R. P., and Dau, T. (**2000**). ''On the role of envelope fluctuation processing in spectral masking,'' J. Acoust. Soc. Am. **108**, 285–296.
Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Glasberg, B. R., and Moore, B. C. J. (**2002**). ''A model of loudness applicable to time-varying sounds,'' J. Audio Eng. Soc. **50**, 331–342.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002a**). ''Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression,'' J. Acoust. Soc. Am. **111**, 2759–2770.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002b**). ''Influence of component phase on the loudness of complex tones,'' Acust. Acta Acust. **88**, 369–377.
Hall, J. L. (**1997**). ''Asymmetry of masking revisited: Generalization of masker and probe bandwidth,'' J. Acoust. Soc. Am. **101**, 1023–1033.
Hellman, R. P. (**1970**). ''Effect of noise bandwidth on the loudness of a 1000-Hz tone,'' J. Acoust. Soc. Am. **48**, 500–504.
Hellman, R. P. (**1972**). ''Asymmetry of masking between noise and tone,'' Percept. Psychophys. **11**, 241–246.
Hellman, R. P., and Zwislocki, J. (**1964**). ''Loudness function of a 1000-cps tone in the presence of a masking noise,'' J. Acoust. Soc. Am. **36**, 1618–1627.
Krumbholz, K., Patterson, R. D., and Nobbe, A. (**2001**). ''Asymmetry of masking between noise and iterated rippled noise: Evidence for time-interval processing in the auditory system,'' J. Acoust. Soc. Am. **110**, 2096–2107.
Lundeen, C., and Small, A. M. (**1984**). ''The influence of temporal cues on the strength of periodicity pitches,'' J. Acoust. Soc. Am. **75**, 1578–1587.
Moore, B. C. J. (**1975**). ''Mechanisms of masking,'' J. Acoust. Soc. Am. **57**, 391–399.
Moore, B. C. J. (**1977**). ''Effects of relative phase of the components on the pitch of three-component complex tones,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
Moore, B. C. J., and Ohgushi, K. (**1993**). ''Audibility of partials in inharmonic complex tones,'' J. Acoust. Soc. Am. **93**, 452–461.
Moore, B. C. J., Alcántara, J. I., and Dau, T. (**1998**). ''Masking patterns for sinusoidal and narrowband noise maskers,'' J. Acoust. Soc. Am. **104**, 1023–1038.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.
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.

Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.
Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.
Richards, V. M. (**1992**). ''The detectability of a tone added to narrow bands of equal-energy noise,'' J. Acoust. Soc. Am. **91**, 3424–3435.
Richards, V. M., and Nekrich, R. D. (**1993**). ''The incorporation of level and level-invariant cues for the detection of a tone added to noise,'' J. Acoust. Soc. Am. **94**, 2560–2574.
Scharf, B. (**1964**). ''Partial masking,'' Acustica **14**, 16–23.
Schroeder, M. R., Atal, B. S., and Hall, J. L. (**1979**). ''Optimizing digital speech coders by exploiting masking properties of the human ear,'' J. Acoust. Soc. Am. **66**, 1647–1652.
Smith, R. L., (**1979**). ''Adaptation, saturation and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Stevens, S. S. (**1972**). ''Perceived level of noise by Mark VII and decibels (E),'' J. Acoust. Soc. Am. **51**, 575–601.
Stevens, S. S., and Guirao, M. (**1967**). ''Loudness functions under inhibition,'' Percept. Psychophys. **2**, 459–465.
Verhey, J. L. (**2002**). ''Modeling the influence of inherent envelope fluctuations in simultaneous masking experiments,'' J. Acoust. Soc. Am. **111**, 1018–1025.
Viemeister, N. F. (**1980**). ''Adaptation of masking,'' in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft U.P., Delft, The Netherlands).
Zwicker, E. (**1958**). ''Über psychologische und methodische Grundlagen der Lautheit,'' Acustica **8**, 237–258.
Zwicker, E. (**1963**). ''Über die Lautheit von ungedrosselten und gedrosselten Schallen,'' Acustica **13**, 194–211.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin).
Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.

Examination of a linear model in an informational masking study

Zhongzhou Tang and Virginia M. Richards
Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104

(Received 24 May 2003; revised 13 February 2003; accepted 14 April 2003)

When multitone maskers are used in a two-interval, forced choice experiment, the amount of masking is larger when the masker is randomly chosen on each presentation interval compared to on each trial (the same masker in the two listening intervals). These conditions are referred to as having within- versus between-trial randomization. If it is assumed that an observer's ultimate detection decision depends on a single decision variable (DV), it is probable that the DV's variance will be substantially larger in the within-trial randomization condition compared to the between-trial randomization condition. The goal of the current experiment is to evaluate the degree to which this stimulus-based change in DV variance can account for the difference in thresholds in the within- versus between-trial randomization conditions. Thresholds are measured for the detection of a tone added to a six-component masker in between- and within-trial randomization conditions. The slopes of the psychometric functions provide an estimate of the variance in the DV for the between- and within-trial randomization conditions. Additionally, a channel model is fitted to the psychophysical results in the within-trial randomization condition. The resulting model is then used to predict the value of the DV for each trial, and ultimately to estimate the proportion of the total variance in the within-trial randomization condition that is attributable to changes in maskers across intervals. The variance of the DV in the between-trial randomization condition accounted for approximately 65% of the total variance in the DV in the within-trial randomization condition. Stimulus-based interval-by-interval masker randomization accounted for approximately 20% of the total variance of the within-trial randomization DV. The remaining 15% of the DV variance in the within-trial randomization condition remained unaccounted for. This result is fairly stable whether the maskers are drawn from a small versus large pool of potential maskers. [DOI: 10.1121/1.1579006]

PACS numbers: 44.66.Dc [MRL]

I. INTRODUCTION

An elevation in signal detection thresholds associated with increased stimulus uncertainty is an example of ''informational masking'' (cf. Watson, 1987). Informational masking is typically contrasted with energetic masking; in the latter case the masker is thought to degrade the representation of the signal at the periphery, whereas in the former case there is little reason to believe the representation of the signal is altered by the masker's presence. As an example of ways in which uncertainty/informational masking can be modulated, if an observer's task is to indicate which of two observation intervals has an added tone, maskers might be drawn at random either across trials or across intervals within a trial. For the former condition masker uncertainty is *between* trials and for the latter condition masker uncertainty is both between trials and *within* the observation intervals of a trial. These two conditions will be referred to as having maskers that are randomized either between or within trials, respectively. Psychophysical results show that thresholds estimated using within-trial randomization can be very large, and are larger than thresholds measured when between-trial randomization is used (cf. Neff and Callaghan, 1988; Neff and Dethlefs, 1995; Wright and Saberi, 1999).

Studies by Wright and Saberi (1999) and Richards *et al.* (2002) show that the change in the physical stimulus across intervals is not sufficient to account for the increase in thresholds for the within- compared to the between-trial randomization conditions. For example, Richards *et al.* (2002) examined the detectability of a 1000-Hz tone added to six tone maskers. In various conditions thresholds were determined when the maskers were drawn at random for each presentation from a pool of masker set sizes ranging from 3 to 24. Because the maskers were drawn with replacement for each presentation, on occasion the maskers presented in the two intervals of a trial were the same. As a result, ''between-trial randomization'' trials occurred in the context of the within-trial randomization condition. Separating the trials in which the maskers happened to be the same versus different in the two intervals indicated slightly superior sensitivity when the maskers were the same, but the modest difference did not reach statistical significance. In contrast, as described above, when data collection is blocked, the difference in thresholds in between- and within-trial randomization conditions is substantial (approximately 10 dB for conditions similar to those used by Richards *et al.*, 2002; see Neff and Callaghan, 1988; Neff and Dethlefs, 1995). Wright and Saberi (1999), who used 10-component maskers, also found large effects of changing the context in which the same trials were presented. Detection thresholds are lower when a block of trials has only two maskers than when the same two maskers are presented on trials interleaved with trials using other maskers (their Figs. 1 and 5, maskers A and B). It is apparent

that the context of across-interval stimulus variability accounts for at least part of the difference in thresholds when maskers are randomly drawn between- versus within-trials.

Acknowledging the limitations of strict stimulus-driven explanations of the difference in thresholds when maskers are varied within compared to between trials, it is nonetheless of interest to examine the role that the change in maskers *per se* contributes to the change in thresholds. To address this issue, it is assumed that the variance of the decision variable (DV) formed in the within-trial randomization condition can be partitioned into three independent pools of variance. One reflects the total variance in the DV formed in the between-trial randomization condition. A second is associated with the change in maskers across observation intervals. The third is a residual variance, presumably reflecting ''context'' effects, including factors associated with efficiency, etc. The first two pools of variance are associated with the stimuli presented on any one trial, and the third is associated with ''context.'' If the third pool of variance is small compared to the other two, the result would indicate that there is relatively little impact of context relative to changes in the physical stimulus on any one trial. This general approach borrows from the work of Lutfi and colleagues (e.g., Lutfi and Doherty, 1994; Oh and Lutfi, 1999).

In order to estimate the total variance in the within-trial and between-trial randomization conditions, psychometric functions are measured in both conditions. The total variance in these conditions is estimated from the squared inverse of the slope of the psychometric function. Next, a model is used to provide an estimate of the variance associated with the change in maskers across intervals. The model is similar to a channel model (cf. Durlach *et al.*, 1986), where it is assumed that the DV is based on the weighted sum of level (in dB) present in different frequency regions. Unlike the Durlach *et al.* model, the current model is not optimal, but instead based on empirically derived linear weights. The model provides a stimulus-based DV, and its variance provides an estimate of the variance associated with the changes in maskers across intervals within a trial.[1] In an effort to ensure that the resulting model estimates are robust, the experiment and fitting procedures are repeated using two different masker ''sets.'' For one set the maskers are drawn at random from a small number of potential maskers and for the other the maskers are drawn at random from a very large set of potential maskers (maskers generated prior to each presentation).

II. METHODS

A. Psychophysical methods

Thresholds were estimated in two conditions; maskers are either randomly drawn between trials or within trials. These two conditions will sometimes be referred to as the between-trial and within-trial randomization conditions. Sometimes the word ''randomization'' is omitted. The methods follow those used by Richards *et al.* (2002), and the results for Obs. 1–3 in the within-trial randomization condition are taken from that experiment. Observers 4–12 are new observers, but the current procedures are essentially identical to those used by Richards *et al.*[2] Briefly, a two-interval, forced-choice procedure is used and the signal level is altered using a 2-down, 1-up staircase (Levitt, 1971). Visual feedback indicating the correctness of the observer's response follows each trial. The ultimate thresholds reported are based on the final 20 threshold estimates. One exception to this rule is for Obs. 3, for whom only 15 threshold replicates contributed to the one of the data sets (set 3C, defined below).

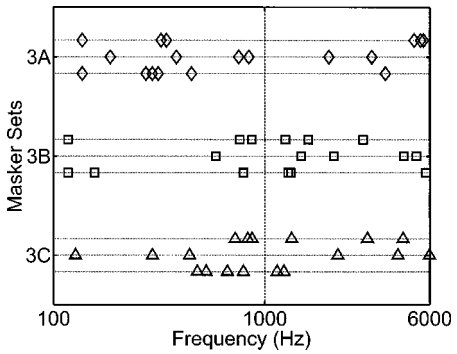


FIG. 1. For the masker sets 3A (top, diamonds), 3B (middle, squares), and 3C (bottom, triangles), symbols show the frequencies of the tonal components that comprise each masker. The *x* axis is frequency. For example, the six tones that comprise one of the three maskers in set 3A (top line) have frequencies of 137, 322, 341, 5049, 5410, and 5586 Hz. The signal frequency is represented using a dotted vertical line.

Maskers are the sum of six tones, whose frequencies could be chosen at random from 100 to 6000 Hz using a uniform distribution on a logarithmic frequency scale. The signal frequency is 1000 Hz, and the masker components are constrained so that they do not fall into a ''protected'' region spanning 929–1080 Hz.

Four distinct masker sets are tested. For one masker set the pool of potential maskers is very large. This set of maskers will be referred to as the ''Ran'' set. For the other masker sets the masker set size is 3, meaning that the maskers are randomly drawn, with replacement, from a set of only three potential maskers. When the set size is 3, thresholds depend on the particular maskers that make up the set. For that reason, three different sets are tested. These sets are referred to as masker set ''3A,'' ''3B,'' and ''3C.'' For these stimulus sets, the maskers were generated prior to the beginning of the experiment, and saved to disk. Figure 1 depicts the maskers that make up masker sets 3A (top, diamonds), 3B (middle, squares), and 3C (bottom, triangles). The abscissa is frequency, and the symbols indicate the frequencies of the components that form each masker. For example, the top three rows of diamonds represent the three maskers that make up set 3A. One of the maskers is set 3A, the one represented in the topmost row, is composed of tones with frequencies of 134, 322, 342, 5049, 5420, and 5586 Hz. The vertical dashed line shows the signal frequency, 1000 Hz. In all four masker sets are tested: 3A, 3B, 3C, and Ran.

The stimuli were digitally generated and presented through two channels of a 16-bit DAC using a sampling rate of 20 000 samples/s, low-pass filtered at 7 kHz using matched filters (KEMO VBF 8), and presented diotically by way of Sennheiser HD410SL headphones. The maskers were presented at 57.8 SPL, 50 dB SPL per tone.

For Obs. 1–8, data were first collected in the within-trial randomization condition. Those observers began with the Ran masker set, and for the remaining masker sets the order of completion was random. Then, data collection began in the between-trial randomization condition, with the four different masker sets (3A, 3B, 3C, and Ran) being run in random order for each observer. It should be noted that for Obs. 1–3 several additional conditions were interspersed between those reported here (see footnote 2).

Observers 9–12 ran a subset of the experiment—only the Ran masker set was tested. Observers 10 and 11 first ran the within-trial condition. Observers 9 and 12 first ran the between-trial condition.

All observers practiced for at least 5 h prior to initial data collection. Whenever observers encountered a new condition or masker set, at least 3 h of practice preceded data collection.

Observers ranged in age from 18 to 29 years and were paid for participation. Only Observer 6 had prior experience in psychoacoustic tasks. Tests were conducted with the observer seated in a double-walled sound-attenuated booth. Observers had thresholds in quiet of 20 dB HL or better for frequencies ranging from 250 to 8000 Hz with the exception of thresholds of 25 dB in one ear at 8000 Hz for Obs. 3 and 12 and 25 dB in the left ear at 6000 Hz for Obs. 2.

B. Computational methods

The analyses followed these steps. First, the total variance of the DV in the within- and between-trial randomization conditions is estimated. The estimates are based on the squared inverse of the slope of the psychometric functions (d' as a function of signal level in dB). Second, a channel model is fitted to the data collection in the within-trial randomization conditions. Third, using the resulting model, an estimate of the variance associated with the across-interval change in maskers is derived. For convenience, this will be referred to as *across-interval variance*. Finally, the across-interval variance is added to the total variance in the between-trial randomization condition to predict the total variance in the within-trial randomization condition. If the psychometric-function-based estimate of total variance in the within-trial randomization condition is similar to the prediction, across-interval variability alone can account for the difference between conditions. If, on the other hand, the psychometric-function-based variance estimate is larger than the predicted variance, it would imply a noise source in the within-trial condition that is not present in the between-trial condition. Such a result might reflect differences in efficiency or processing strategies, etc. Next, the computational steps are described in more detail.

For each masker set (3A, 3B, 3C, and Ran) and each observer two psychometric functions are fitted: one for the within-trial condition and one for the between-trial condition. It is assumed that the function relating d' and signal strength in dB is linear.[3] The data are from adaptive staircase runs. For each signal level encountered the data are included in the psychometric function fit if two criteria are achieved: (a) at least five trials contribute to each percent correct, and (b) percent correct is above 40% and below 98%. The d' values are estimated from the percent-correct scores using a numerical approximation to the *p-to-z* transform (Abramowitz and Stegun, 1964). The least-squares linear fit minimized a chi-square in which the error at each signal level was weighted by an estimate of the standard deviation (square root of the number of trials) and then squared.

Importantly, the psychometric functions for the within-trial and between-trial conditions are not derived independently. Rather, the two linear functions are forced to share a single x intercept. This step encompasses the fact that in the between- and within-trial conditions d' will be zero at approximately the same signal strength. There appears to be little cost in incorporating the common-x-intercept assumption; when the within- and between-trial condition psychometric functions are fitted separately, the resulting intercepts are not statistically different from one another. In summary, then, the data from the within- and between-trial conditions are fitted using three parameters: two slope parameters and one x-intercept parameter.

The slopes of the lines fitted to the psychophysical data in the between- and within-trial conditions provide estimates of the total variance in the two conditions (i.e., the standard deviation is estimated from the inverse of the slope). The next step involves the estimate of the across-interval variance for each of the masker sets (3A, 3B, 3C, and Ran). For each stimulus tested, the spectral components are sorted into 28 bins of equal bandwidth on a logarithmic scale ranging in frequency from 100 to 6000 Hz. At 1000 Hz the bandwidth is 160 Hz. The stimulus strength in each bin is the value of the bin in dB SPL minus the level at which performance is at chance (the x intercept from the psychometric functions, in dB SPL). Then, relative weights are estimated using a linear model fitted to the trial-by-trial responses. The assumption implicit in this approach is that observer's responses depend on the linear combination of level in dB across 28 independent and nonoverlapping channels.

Figure 2 shows the computation steps for the Ran masker set. The top two panels plot the spectra of the stimuli presented in the two listening intervals—the 1000-Hz signal is present in the first interval. The third panel shows the difference between spectra; interval 1 minus interval 2. The fourth panel plots the weighting function derived for the observer, Obs. 11 in this case. The bottom panel shows the weighted difference. The decision variable, DV, is the sum of these values, in dB. Because the model is linear, this procedure is equivalent to the case in which the power spectra in each interval are weighted and summed, and the resulting random variables differenced. The variance of the DVs across all trials is an estimate of the across-interval variance. This value is a numerical estimate of the uncertainty associated with changes in maskers across intervals.

For this model, an optimal strategy is one in which all nonsignal channels have equal weights and the signal-channel weight is nonzero. For example, the nonsignal weights might be zero. When the masker set size is 3, there are not significant degrees of freedom to estimate relative weights at many frequencies (e.g., 28 bins are used when the Ran masker set is tested). For these masker sets, QR factorization is used to estimate the relative weights for an

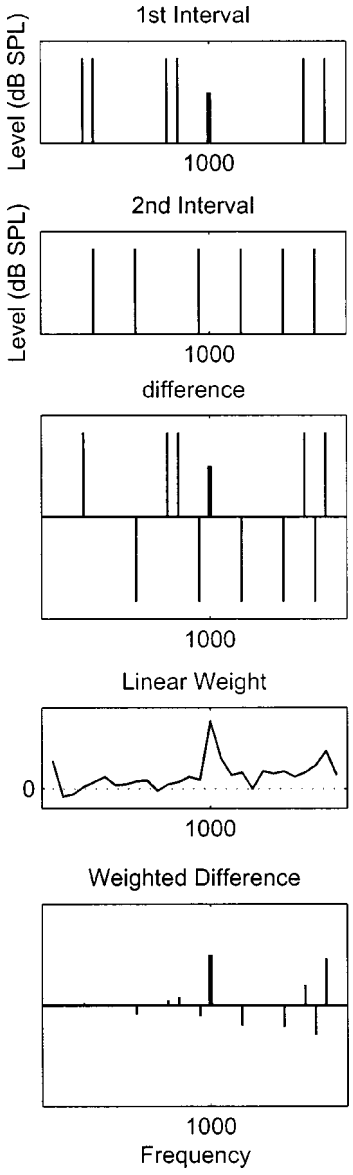


FIG. 2. The top two panels show the spectra for stimuli presented in the first and second intervals, respectively. The first interval has the added 1000-Hz signal. The third panel shows the differenced power spectra and the fourth panel shows the relative weights estimated for Obs. 11. The final panel shows the product of the difference spectra and the weighting function. The decision variable is the sum of these values.

n-parameter linear model (cf. Strang, 1980). In principle, *n* might take on a value other than 3, but for the current data set, $n=3$. Additional details of the model fits are described in Richards *et al.* (2002).

Finally, adding the estimate of the across-interval variance to the total variance in the between-trial randomization condition provides a prediction of the total variance in the within-trial randomization condition. Below, standard deviations, not variances, are presented. The reliability of estimated and predicted total standard deviations of the DV in the within-trial randomization conditions are estimated using resampling methods (e.g., Efron and Tibshirani, 1993). Bootstrap data sets are redrawn from the original responses, and the complete model fitting procedure repeated 1000 times. The standard deviation of the 1000 bootstrap estimates provides an estimate of the standard error of the mean.

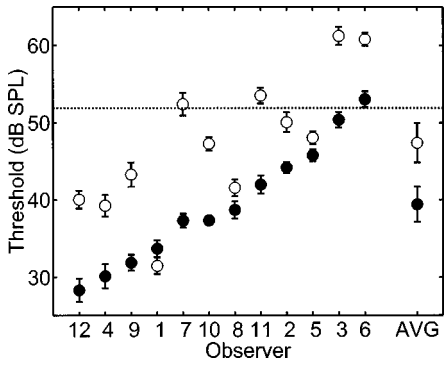


FIG. 3. Thresholds in dB SPL for the Ran masker set are plotted for individual observers and the average (AVG, far right). Filled circles are for the between-trial randomization condition and unfilled circles are for the within-trial randomization condition. The observers are ordered according to their thresholds in the between-trial randomization condition. For the individual data, error bars indicate standard errors of the mean across 20 threshold replicates, and for the averaged data the error bars show standard errors of the mean across observers.

In all, four key assumptions are made in the modeling effort. First, it is assumed that when the maskers are changed across intervals, the variance in the DV is increased. In the current model, this reflects the integration of energy at frequencies other than the signal frequency. Second, it is assumed that the slope of the psychometric function is linear when d' is plotted as a function of dB, and that the inverse of the slope of the psychometric function is equal to the standard deviation of the DV. Third, it is assumed that energy is integrated across frequencies other than the signal frequency and that the combination can be treated as a weighted linear combination of dB values. Fourth, it is assumed that performance falls to chance levels ($d'=0$) at the same signal level regardless of whether maskers are varied across vs between trials.

III. RESULTS AND DISCUSSION

Figure 3 shows thresholds for the individual observers and the average across observers (far right, AVG) for the Ran masker set. Thresholds for the between- and within-trial randomization conditions are shown using filled and unfilled circles, respectively. Thresholds are in dB SPL. The observers are ordered by threshold in the between-trial randomization condition; the *x* axis indicates observer identity. The dotted horizontal line indicates the signal level needed to elevate the signal-plus-masker stimulus 1 dB above the masker-alone stimulus, a rough estimate of the signal level at which observers can accomplish the task on the basis of overall changes in intensity.

Consistent with past reports (cf. Neff and Dethlefs, 1995), there are substantial individual differences in thresholds. Even so, only Obs. 1's results indicate a lower threshold in the within- than the between-trial randomization conditions (Obs. 1 is the fourth observer when rank-ordered). On average thresholds in the within-trial condition are 8 dB higher than in the between-trial condition. For similar conditions, Neff and Callaghan (1988) report approximately a 10-dB difference between the between- and within-trial conditions (their fixed and random conditions, respectively).

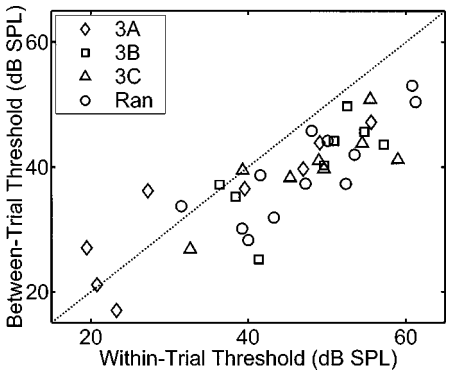


FIG. 4. Thresholds in the between-trial condition are plotted against thresholds in the within-trial condition. The diagonal line indicates points of equal threshold. Thresholds for sets 3A, 3B, 3C, and Ran are indicated using diamonds, squares, triangles, and circles, respectively.

Figure 4 plots thresholds for the Ran, 3A, 3B, and 3C stimulus sets on one plot. Thresholds in the between-trial condition are plotted as a function of thresholds in the within-trial condition. Thresholds for the Ran stimulus set are plotted using circles; thresholds for stimulus sets 3A, 3B, and 3C are plotted using diamonds, squares, and triangles, respectively. The data points tend to fall below the equal-threshold diagonal, indicating thresholds tend to be higher in the within- than the between-trial randomization condition.

Thresholds for set 3A (diamonds) are generally lower than the other masker sets. Considering the distribution of masker components shown in Fig. 1, one difference between the maskers in stimulus set 3A and the others is that the maskers tend not to have components with energy in the region just above the signal frequency, 1000 Hz. With so few maskers tested, however, it is difficult to determine whether this accounts for the difference in masking. Consistent with work reported by Neff and Dethlefs (1995), across masker sets thresholds grow more quickly in the within-trial conditions than in the between-trial conditions. A repeated-measures ANOVA applied to the data of the eight observers who completed all four stimulus sets (3A, 3B, 3C, and Ran) revealed a main effect of stimulus set [$F(3,21)=10.8$, $p<0.001$] and a main effect of within- vs between-trial randomization conditions [$F(1,7)=22.4$, $p<0.001$]. The interaction approached but did not reach significance ($p<0.1$). This modest evidence for an interaction reflects the fact that thresholds in the within- and between-trial randomization conditions tend to be more similar for stimulus sets 3A than the other stimulus sets.

The first two data columns of Table I list the slopes of the psychometric functions fitted to the data obtained in the between- and within-trial randomization conditions. The table lists the slopes for the individual observers and for the different stimulus sets. The single x intercept fitted to the functions is shown in the third column. The percent of variance in the psychometric functions (accumulated across between- and within-trial data sets) accounted for using linear fits is indicated in the fourth column. For Obs. 1–8 the fitted parameters are indicated for four stimulus sets; Ran, 3A, 3B, and 3C. For Obs. 9–12, data were collected only using the Ran masker set.

TABLE I. Slopes for between- (slope B) and within-trial (slope W) masker randomization conditions, the joint intercept, and the percent of variance in the psychometric functions accounted for by the linear fits are listed for individual observers in different conditions.

Observer	Condition	Slope B	Slope W	Intercept	% Var
Obs. 1	3A	0.107	0.103	12.1	79
	3B	0.083	0.061	23.6	61
	3C	0.036	0.032	12.0	50
	Ran	0.056	0.068	17.1	81
Obs. 2	3A	0.056	0.042	24.2	55
	3B	0.081	0.057	33.3	62
	3C	0.092	0.048	30.5	68
	Ran	0.104	0.067	35.6	73
Obs. 3	3A	0.107	0.057	38.1	65
	3B	0.035	0.020	19.2	40
	3C	0.062	0.028	25.8	56
	Ran	0.068	0.035	36.7	67
Obs. 4	3A	0.126	0.071	9.5	82
	3B	0.085	0.043	13.8	81
	3C	0.075	0.053	14.1	73
	Ran	0.046	0.039	10.4	52
Obs. 5	3A	0.142	0.098	30.9	83
	3B	0.129	0.057	34.0	70
	3C	0.101	0.047	34.6	68
	Ran	0.092	0.063	35.7	74
Obs. 6	3A	0.086	0.056	32.0	66
	3B	0.078	0.066	38.0	66
	3C	0.081	0.064	39.6	70
	Ran	0.079	0.051	41.2	57
Obs. 7	3A	0.059	0.093	9.7	80
	3B	0.032	0.036	4.2	30
	3C	0.041	0.031	14.8	35
	Ran	0.072	0.030	23.9	59
Obs. 8	3A	0.046	0.089	16.7	60
	3B	0.056	0.029	27.1	49
	3C	0.074	0.048	26.0	64
	Ran	0.055	0.051	22.0	69
Obs. 9	Ran	0.036	0.020	13.0	62
Obs. 10	Ran	0.067	0.031	28.0	62
Obs. 11	Ran	0.037	0.023	25.3	51
Obs. 12	Ran	0.039	0.024	9.6	55
	Mean	0.076	0.054	24.6	64
	Sem	0.005	0.004	1.9	2

The psychometric functions tend to be steeper in the between- than in the within-trial randomization conditions; the opposite is observed on only four occasions. Wright and Saberi (1999) also report that on average psychometric function slopes are steeper when the maskers are randomly chosen between than within trials. Notably, the slopes of their psychometric functions are approximately 1.5 times as steep as those shown in Table I. We have no sense of which of the many experimental/observer differences led to this sizable change in slope. Estimating the total DV variance associated with each condition as the squared inverse of the slope of the psychometric function, there is more total variance in the within- than the between-trial randomization conditions.

Figure 5 shows the estimates of the relative weights as a function of frequency in the within-trial randomization condition when the maskers were randomly drawn (Ran condition). These results are for Obs. 4–12. For Obs. 1–3 the results are similar to these, and essentially identical to patterns previously published for those observers (Richards *et al.*, 2002).[4] In general, in the Ran condition the weight

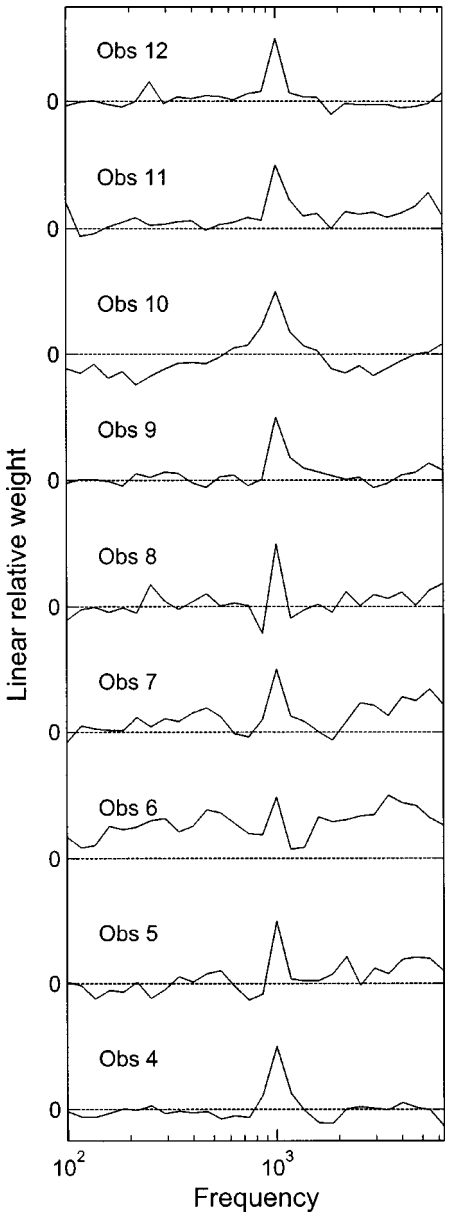


FIG. 5. Relative weights are plotted as a function of frequency for the Ran condition for Obs. 4–12.

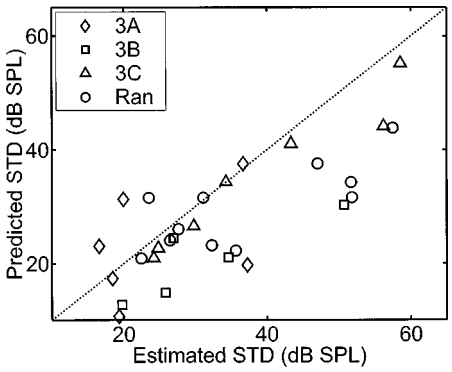


FIG. 6. The predicted standard deviations in the within-trial randomization condition are plotted as a function of the psychometric-function-based estimates of the total standard deviations. Results for sets 3A, 3B, 3C, and Ran are plotted using diamonds, squares, triangles, and circles, respectively.

associated with the signal frequency is larger than the other weights, and the nonsignal weights are close to zero. In some instances the higher-frequency weights tend to be larger than the lower-frequency weights (e.g., Obs. 5 and 7). For Obs. 6 in the Ran condition the weighting function is substantially different from the others; across frequency the weights are positive, although variable. There are several potential interpretations of such a pattern of weights. For example, the observer's decision might randomly rely on different frequency regions on different trials. Another potential interpretation is that the observer depends on energy present at all frequencies (summed on a dB scale), akin to an energy detector. This latter interpretation is consistent with the fact that her Ran thresholds exceeded the threshold predicted based on loudness (Fig. 3, threshold above the dotted line). However, Obs. 6 is not the only observer whose within-condition thresholds are higher than the loudness bound; four of the 12 observers failed to best that bound. For the other three observers, the pattern of weights is not consistently positive for all frequencies.

Given the relative weights, the magnitude of the across-interval variance is estimated for each observer using the exact stimuli s/he encountered in the experiment. This variance is added to the estimate of the total variance in the between-trial randomization condition (the squared inverse of the slope of the psychometric function) to generate a prediction of the total variance for the within-trial randomization condition. For Obs. 6 in the Ran condition, the psychometric-function-based estimate of total variance in the within-trial condition minus the across-interval variance is negative. This is a clear violation of the linear model, possibly owing to errors in parameter estimation, and thus her Ran results are not included in the following.

Figure 6 compares the standard deviation in the within-trial randomization condition estimated using the obtained psychometric functions (abscissa) and the predicted standard deviation generated using the between-trial randomization data and the estimate of the across-interval variance (ordinate). Results in the Ran, 3A, 3B, and 3C conditions are plotted using circles, diamonds, squares, and triangles, respectively. The diagonal line indicates the set of points for which the prediction is perfect. Most of the data points fall below the diagonal, indicating that the sum of the total variance in the between-trial randomization condition plus the across-interval variance falls short of predicting the total variance in the within-trial randomization condition.

It is of interest to consider the relative contribution of the different sources of variance in the within-trial randomization condition. Averaging across observers and masker sets, the breakdown is as follows. The total variance in the between-trial randomization condition is approximately 65% of the total variance in the within-trial randomization condition. The variance associated with across-interval changes in maskers is approximately 20% of the total variance in the within-trial randomization condition, and 15% of the total variance in the within-trial condition remains unaccounted for. Thus, approximately 85% of the total variance in the within-trial condition can be accounted for by summing the total variance in the between-trial conditions and the vari-

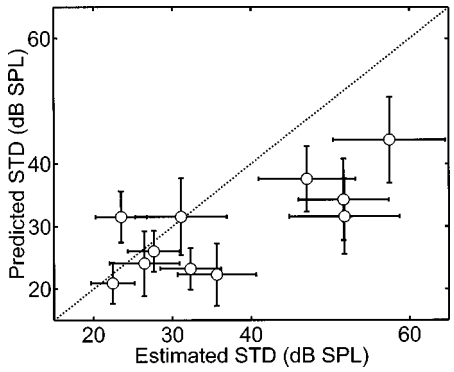


FIG. 7. As Fig. 6, except the results are only for the Ran condition. Error bars indicate an estimate of the standard error of the mean of the estimates.

ance associated with across-interval changes in the masker. This breakdown is fairly consistent across the Ran, 3A, 3B, and 3C masker sets.

Next, consider the reliability of the estimates, and the primary sources of error in parameter estimates. The predicted and estimated standard deviations of the DV in the Ran condition shown in Fig. 6 are plotted in Fig. 7. Error bars indicating an estimate of the standard error of the mean are also shown (Efron and Tibshirani, 1993). It is apparent that there is substantial error in estimation. The primary contributing factor is that shallow psychometric functions generate unreliable slope estimates. By comparison the variability in the estimate of the relative weights is minimal.

To summarize, whether the masker set size is small or large, changing from between- to within-trial masker randomization leads to an approximate 35 percent increase in the variance of the DV. Provided the assumptions associated with the modeling effort are accurate, the increment in DV variance leads to a 60/40 split between the variance associated with across-interval changes in the maskers and a ''residual'' variance, respectively. The source of the residual variance is not specified; it might reflect changes in efficiency, an impact of the large difference of the percepts associated with different maskers that is not captured by the model tested here, internal noise associated with generalized uncertainty, etc.

IV. SUMMARY AND CONCLUSIONS

Consistent with past results, the current experiment indicates large individual differences in informational masking, even when the number of maskers tested is small. Whether the maskers are varied within trials or between trials, the psychometric functions are fit reasonably well with linear functions in d' vs signal level (in dB). The slopes are generally shallower when masker randomization is within trials compared to between trials. This result holds whether the maskers are drawn from a large (Ran) or small (3A, 3B, 3C) pool of potential maskers. It is estimated that the total variance in the between-trial randomization condition is 65% of that in the within-trial randomization condition. Across-interval variation in maskers accounts for roughly 60% of the remaining 35% of the total variance present in the within-trial randomization condition.

ACKNOWLEDGMENTS

This work was supported by Grants RO1 DC 02012 from the National Institutes of Health. Dr. Gerald D. Kidd, Jr., Dr. Robert A. Lutfi, and an anonymous reviewer provided very helpful comments/direction on an earlier version of this manuscript.

[1]Because the weights associated with the channel model are based on a limited (although large) number of trials, there is the possibility that the estimate of the weights will depend on the magnitude of residual variance. There is no reason to expect such an error to lead to biases in the relative weights.

[2]In the Richards *et al.* (2002) paper, the current Obs. 1–3 were identified as Obs. 1, 3, and 4, respectively. There are only two differences in procedures for Obs. 1–3 compared to Obs. 4–12: the conditions were run in a different order and Obs. 1–3 ran many more conditions than Obs. 4–12. The additional conditions Obs. 1–3 ran included thresholds when the masker set size was 6, 12, and 24. In addition, due to software changes, the time between intervals was inadvertently changed—the interstimulus interval was 400 ms for Obs. 1–8 and 600 ms for Obs. 9–12.

[3]Dr. Robert A. Lutfi and Christine R. Mason independently suggested we consider fitting the psychometric functions using a form in which a lapse parameter may be introduced. This suggestion reflects their observation that observers' performance levels sometimes asymptote to percent-correct levels below 100 percent correct. If the psychometric function asymptotes at a value less than 100 percent correct, fitting the functions using a line on d' coordinates will lead to an underestimate of the slope, and thus an overestimate of the total variance. When the percent-correct data were fitted using a logistic function with a lapse parameter, the results did not indicate substantive differences in the relation between the slopes in the within- and between-trial conditions.

[4]The data for Obs. 1–3 of the current paper were also used to generate the functions in Richards *et al.* (2002). The current functions are slightly different because the current fits used the increment in level relative to an estimated x intercept, not just dB SPL. This leads to a modest increase in the relative magnitude of the signal weight compared to the others, but the quality of fit is unaltered.

Abramowitz, M., and Stegun, I. A. (**1964**). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York).

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.

Efron, B., and Tibshirani, R. J. (**1993**). *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability 57 (Chapman and Hall, New York).

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 167–177.

Lutfi, R. A., and Doherty, K. A. (**1994**). ''Effect of component-relative-entropy on the discrimination of simultaneous tone complexes,'' J. Acoust. Soc. Am. **96**, 3443–3450.

Neff, D. L., and Callaghan, B. P. (**1988**). ''Effective properties of multicomponent simultaneous maskers under conditions of uncertainty,'' J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Dethlefs, D. L. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.

Oh, E. L., and Lutfi, R. A. (**1999**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **106**, 3521–3528.

Richards, V. M., Tang, Z., and Kidd, Jr., G. D. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.

Strang, G. (**1980**). *Linear Algebra and its Applications* (Academic, New York).

Watson, C. S. (**1987**). ''Uncertainty, informational masking, and the capacity of immediate auditory memory,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).

Wright, B. A., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.

Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity[a)]

Nathaniel I. Durlach,[b)] Christine R. Mason, Barbara G. Shinn-Cunningham, Tanya L. Arbogast, H. Steven Colburn, and Gerald Kidd, Jr.
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 27 September 2002; revised 20 March 2003; accepted 8 April 2003)

Previous work has indicated that target-masker similarity, as well as stimulus uncertainty, influences the amount of informational masking that occurs in detection, discrimination, and recognition tasks. In each of five experiments reported in this paper, the detection threshold for a tonal target in random multitone maskers presented simultaneously with the target tone was measured for two conditions using the same set of five listeners. In one condition, the target was constructed to be ''similar'' (S) to the masker; in the other condition, it was constructed to be ''dissimilar'' (D) to the masker. The specific masker varied across experiments, but was constant for the two conditions. Target-masker similarity varied in dimensions such as duration, perceived location, direction of frequency glide, and spectro-temporal coherence. Group-mean results show large decreases in the amount of masking for the D condition relative to the S condition. In addition, individual differences (a hallmark of informational masking) are found to be much greater in the S condition than in the D condition. Furthermore, listener vulnerability to informational masking is found to be consistent to at least a moderate degree across experiments. [DOI: 10.1121/1.1577562]

PACS numbers: 43.66.Dc, 43.66.Lj [MRL]

I. INTRODUCTION

There is considerable room for argument about how best to define informational masking or even whether the term ''informational masking'' is useful. Nevertheless, it is generally agreed that informational masking is distinct from energetic masking, where energetic masking is defined as the masking that results from competition between target and masker at the periphery of the auditory system (i.e., at the level of the basilar membrane or auditory nerve). Consistent with this characterization is the idea that informational masking reflects vulnerability of certain central portions of the auditory processing system (e.g., related to attentional phenomena). Furthermore, it has been amply demonstrated that substantial amounts of informational masking can be created through the introduction of uncertainty in the acoustic stimulus. In fact, some investigators have used the effects of uncertainty to define informational masking (e.g., Watson and Kelly, 1981; Neff, 1995; Oh and Lutfi, 2000). Although it has not been shown that stimulus uncertainty is either a necessary or a sufficient condition to produce nonenergetic masking, there is no doubt that stimulus uncertainty can produce large amounts of such masking under a wide variety of conditions. Further comments on some of these conceptual issues are available in Durlach *et al.* (2003a).

In this paper, we report the results of a series of detection experiments (involving tonal targets and random multitone maskers presented simultaneously with the target) designed to demonstrate that informational masking resulting from stimulus uncertainty can be substantially reduced by introducing target-masker dissimilarity. To the extent that this is in fact the case, one can conclude that an adequate theory of informational masking must take account of target-masker similarity as well as uncertainty. The experimental results reported in this paper add to previously reported informational masking experiments concerned with target-masker similarity not only by providing additional data, but by exploring additional dimensions of similarity and by employing the same set of listeners in all of the experiments. Inasmuch as intersubject variability is known to be extraordinarily large in informational masking experiments, comparing the results for individual listeners across several different experiments can provide important additional insight into the nature of informational masking.

II. BACKGROUND

Previous research on informational masking includes studies of both simultaneous and sequential masking for discrimination and recognition as well as for detection, for a wide variety of signal types ranging from simple tonal stimuli to running speech. The following comments focus on the empirical work on informational masking for simultaneous nonspeech stimuli in which the target is a fixed-frequency tone and the masker is a multitone complex whose frequency components are varied randomly from presentation to presentation or trial to trial (e.g., Neff and Green, 1987; Neff and Callaghan, 1987; Neff *et al.*, 1993; Neff, 1995; Neff and Dethlefs, 1995; Kidd *et al.*, 1994; Oh and Lutfi, 1998; Wright and Saberi, 1999; Richards *et al.*, 2002). The spacing of the masker components is restricted in such a way that relatively little masker energy occurs in the fre-

[a)]Portions of this work were presented at the Acoustical Society of America meeting in Pittsburgh, PA in June 2002 [Mason *et al.*, J. Acoust. Soc. Am. **111**, 2470(A) (2002)].

[b)]Electronic mail: durlach@mit.edu

quency region around the signal tone (called the ''protected region''). Because the amount of energetic masking decreases with the size of the protected region (cf. Neff *et al.*, 1993), many of the experiments designed to focus on informational masking use protected regions that are equal to or greater than the ''critical band'' around the given target frequency. Also, because in many cases both energetic and informational masking are expected to occur at least to some degree, attention is given to how much masking is energetic and how much is informational, and to how the two types of masking interact (e.g., Lutfi, 1990). The number of tonal components in the masker, and the frequency range of those components, as well as the extent to which the components are randomized in amplitude and frequency, varies with the experiment. Also, the relative effect of randomizing the spectrum of the masker between intervals and between trials in two-interval paradigms has been examined (cf. Neff and Green, 1987; Neff, 1995; Neff and Dethlefs, 1995; Wright and Saberi, 1999; Richards *et al.*, 2002).

The effect of randomizing the spectrum of the masker can be exceedingly large. However, the results of such experiments are strongly listener dependent. Whereas some listeners, occasionally referred to as ''holistic'' or ''synthetic'' listeners, evidence very large effects of the uncertainty in the multitone masker, other listeners, often referred to as ''analytic'' listeners, show hardly any effect at all (Espinoza-Varas and Watson, 1989; Neff and Dethlefs, 1995; Lutfi *et al.*, 2003). Moreover, it appears that the variation in the size of this effect arises primarily from variation in the masked threshold for the uncertain-masker case rather than for the certain-masker case (which is often broadband noise). Questions of current interest in this area include: To what extent does a listener's ability to resist informational masking vary with the experimental task? What other differences among listeners correlate with this ability? How much can this ability be enhanced by training? According to a recent study by Oxenham *et al.* (2003), there is a significant positive correlation between resistance to informational masking and musical training.

Despite the large amount of data on informational masking that has become available over the past few years, there have been only a few attempts to model informational masking. Currently there is no model that satisfactorily accounts for all of the empirical results, even when limited to the body of work on detecting a target tone in a simultaneous random multitone masker discussed above. The most extensive effort to date is the CoRE (component relative entropy) model proposed by Lutfi (1993). Oh and Lutfi (1998) have shown that the CoRE model, which uses the weighted outputs (mean levels and variances) of a set of peripheral filters in addition to a variable bandwidth ''attentional'' filter, can predict the variation in threshold with number of masker tones (as originally found by Neff and Green, 1987) with considerable accuracy. In other cases, however, such as the detection threshold for an inharmonic tone embedded in a randomized harmonic multitone masker, the model is less successful (Oh and Lutfi, 2000). In distinct but related efforts, both Wright and Saberi (1999) and Richards *et al.* (2002) have interpreted informational masking data in terms of channel-weighting analyses.

Apart from the modeling work noted above, which is focused primarily on uncertainty in the stimulus combined with channel weights, the main theoretical notions that have been proposed to help understand informational masking phenomena concern the perceptual grouping or segregation of target and masker (Leek *et al.*, 1991; Kidd *et al.*, 1994; Neff, 1995; Oh and Lutfi, 2000). At a crude intuitive level, informational masking occurs because the listener finds it difficult to focus attention on the target in the presence of a distracting or confusing masker. Although uncertainty is clearly relevant to this phenomenon, so is the extent to which the target ''sounds like'' the masker and is grouped with the masker. In the words of Leek *et al.* (1991, pp. 205–206), ''Informational masking is broadly defined as a degradation of auditory detection or discrimination of a signal embedded in a context of other similar sounds'' and ''A target that is sufficiently different from the surrounding tones along some acoustic dimension will be heard with increased precision.'' Thus, in addition to uncertainty, similarity, which is well known to be a factor in the extent to which auditory objects may be grouped into a single auditory image or segregated into separate images (e.g., Bregman, 1990), has also been considered as an important factor in informational masking.[1]

In the studies by Kidd *et al.* (1994) and Neff (1995), informational masking was reduced by decreasing the similarity between target and masker in a variety of dimensions including spectro-temporal pattern, duration, and interaural (i.e., spatial) relationship. In the Neff study (1995, p. 1910), the purpose was ''to increase the perceptual differences between the signal and the sinusoidal masker components and thus facilitate hearing out the signal from the tonal complex.'' In the Kidd *et al.* study (1994, p. 3475), the stimulus manipulations were chosen so that ''they produced the subjective impression that the signal and masker were perceptually segregated into different auditory 'objects' or 'images.' '' The results of both of these studies are considered along with our own results in Secs. IV and V. In the paper by Oh and Lutfi (2000) on harmonicity mentioned above, the authors state that ''large elevations in threshold are often attributed to the lack of any predictable structure in the masker that would allow listeners to segregate the single spectral component belonging to the signal from the collection of the components belonging to the masker'' (p. 706). Accordingly, they hypothesized that a harmonic masker should produce less masking when the target is not one of the harmonic components. Their results were consistent with this conjecture about the role of perceptual segregation in reducing informational masking (but could not be accounted for by the CoRE model).

In general, it seems clear that the amount of informational masking cannot be predicted solely on the basis of uncertainty (even when the computation of uncertainty goes beyond consideration of the uncertainty in the stimulus waveforms) and that target-masker similarity and the phenomena of grouping and segregation must also be considered. Although the extent to which the components of a complex acoustic stimulus are grouped and segregated into

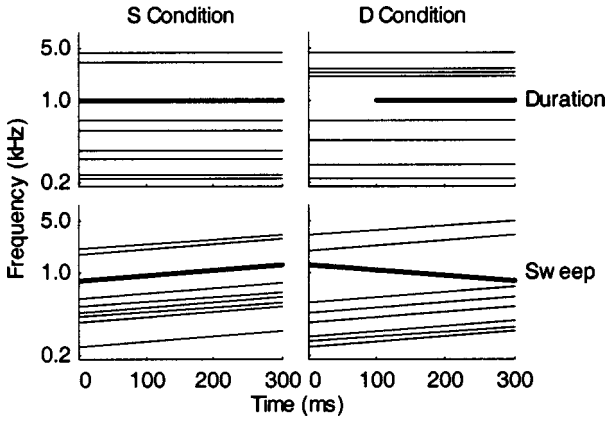


FIG. 1. Schematic illustration of the Duration experiment (top panels) and the Sweep experiment (bottom panels). All four panels show the case when the target is present (bolder lines). The two panels on the left illustrate the S condition and the two on the right the D condition. In the Duration experiment, the S and D conditions differ by target duration. In the Sweep experiment, the S and D conditions differ by the target sweep direction.

distinct images is difficult to quantify, there is some hope that eventually one or more metrics of target-masker similarity can prove useful in predicting the amount of informational masking that occurs. It should also be noted that target-masker similarity appears to be important in a wide range of complex auditory detection and recognition tasks. For example, there is substantial evidence that target-masker similarity plays a major role in speech reception tasks: informational masking tends to increase as the masker goes from noise to speech to same-sex talker to same talker (e.g., Freyman *et al.*, 1999, 2001; Brungart, 2001; Brungart *et al.*, 2001; Arbogast *et al.*, 2002).[2] Furthermore, a recent study by Kidd *et al.* (2002) provides support for the proposition that target-masker similarity affects informational masking for nonspeech pattern recognition. Finally, it should be noted that similarity is a well-known factor in the degree to which stimuli interfere with or mask each other in sequential as well as simultaneous masking [for extensive work on temporal patterns and sequential masking, see the work by Watson and his colleagues as exemplified in Watson *et al.* (1976), Watson and Kelly (1981), Watson (1987), and Espinoza-Varas and Watson (1989)] and in sensory channels other than audition [see, for example, Turvey (1973) for a consideration of pattern masking in vision].

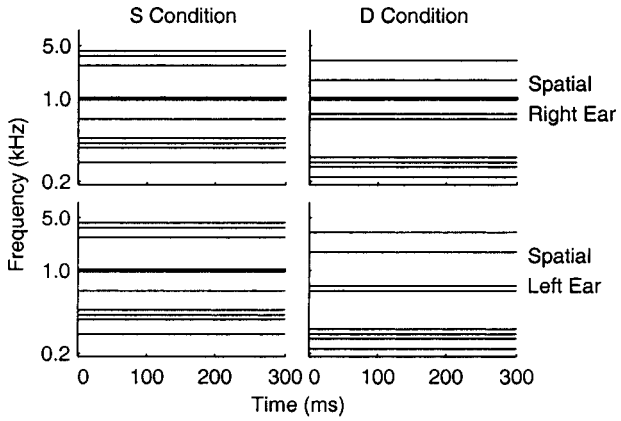


FIG. 2. Schematic illustration of the Spatial experiment. All panels show the case when the target is present (bolder lines). The top panels are the stimuli presented to the right ear and the bottom panels are the stimuli for the left ear. In the S condition (left panels), both the target and masker are presented diotically. In the D condition (right panels), the masker is presented diotically and the target monotically.

The purpose of the present study was to examine informational masking, and release from informational masking, for conditions in which target-masker similarity was varied while masker uncertainty remained unchanged. This work is thus a relatively direct extension of the previous investigations by Kidd *et al.* (1994), Neff (1995), and Oh and Lutfi (2000). In addition to testing new conditions in which target-masker similarity is varied, all of the listeners participated in all conditions of each experiment, thus allowing a determination of the consistency of listener performance across a diverse set of masking conditions. The expectation is that this data set will prove useful in future efforts to model informational masking that take into account the similarity and/or perceptual grouping and segregation of sounds.

III. EXPERIMENTS

A. Overview

Schematic diagrams of the experiments performed are shown in Figs. 1–3. In all cases, S is used to denote ''target and masker similar'' and D to denote ''target and masker dissimilar.'' The following paragraphs contain brief descriptions of the experiments performed (further details about these experiments are given in Sec. III B). In each case, a multitone masker with a high degree of frequency uncertainty is used. The distinction between the S and D conditions is made by changes to the target only and therefore involves no change in the masker uncertainty. In each case, it is intuitively obvious that the masked threshold in the D condition should be lower than in the S condition, despite the fact that such a result cannot reasonably be predicted on the basis of either stimulus energy or stimulus uncertainty (e.g., only in the fifth experiment is there any decrease in stimulus uncertainty in going from the S condition to the D condition, and even in that case it seems very unlikely that uncertainty rather than target-masker similarity is the relevant issue).

Although all the experiments performed were alike in the sense that the uncertainty in the stimulus consisted of frequency uncertainty in the masker, they differed with respect to the parameters used to manipulate the degree of similarity between target and masker. The first experiment made use of duration, the second of direction of frequency sweep, the third of interaural parameters influencing spatial perceptual characteristics, and the fourth and fifth of parameters influencing grouping and streaming perceptual characteristics. Taken all together, it is believed that the array of stimulus parameters (and the subjective counterparts of these parameters) manipulated constitutes a sample that is sufficiently broad to enable one to draw general conclusions about the interaction of similarity and uncertainty with reasonable safety.

1. The shortened-target-duration experiment (Duration)

As shown in the top two panels of Fig. 1, the S and D conditions differ by target duration only: the target in the D

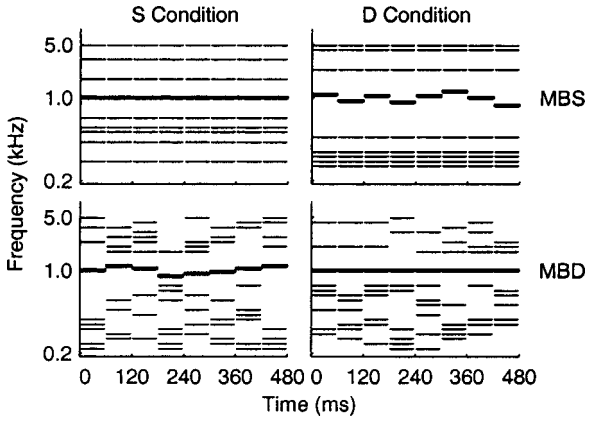


FIG. 3. Schematic illustration of the MBS and MBD experiments. All graphs show the case when the target is present (bolder lines). The top two panels show the MBS paradigm. In this case, the masker was always held fixed from burst to burst. In the S condition (top left), the target was also held fixed; in the D condition (top right), it was jittered in frequency from burst to burst. The bottom two panels show the MBD paradigm. In this case the masker was always jittered. In the S condition (bottom left), the target was also jittered; in the D condition (bottom right), the target was held fixed.

condition is shorter than the target in the S condition (i.e., the target is turned on after the onset of the masker). This experiment is similar to one performed by Neff (1995).

2. The reversed-frequency-sweep experiment (Sweep)

As shown in the bottom two panels of Fig. 1, the masker tones are all upward frequency glides. In the S condition, the target is a glide with the same extent and direction as the masker components. In the D condition, the target glide is in the opposite direction from the masker components.

3. The separate-spatial-channels experiment (Spatial)

Figure 2 illustrates the third experiment. The S condition consists of a diotic multitone masker with a diotic tonal target. In the D condition, the target is simply removed from one ear so that the masker is diotic and the target is monotic. This experiment is similar to a condition in the Kidd *et al.* (1994) paper and is similar in intent to a binaural condition in Neff (1995).

4. The jittered-target-frequency experiment (MBS)

The top two panels of Fig. 3 illustrate the fourth experiment. In both the S and D conditions, the masker and target consist of multiple-burst stimuli. Whereas the frequencies of the masker components are always held fixed from burst to burst [referred to as multiple-bursts-same, or MBS, as in Kidd *et al.* (1994)], the frequency of the target tone is fixed in the S condition but jittered in the D condition.

5. The constant-target-frequency experiment (MBD)

The bottom two panels of Fig. 3 illustrate the fifth experiment. As in the fourth experiment, multiple-burst stimuli are employed in both the S and D conditions. However, in this experiment, whereas the frequencies of the masker are now always randomized from burst to burst [referred to as multiple-bursts-different, or MBD, as in Kidd *et al.* (1994)], the frequency of the target tone is now jittered in the S condition but held fixed in the D condition.

B. Methods

1. Listeners

Five university students between the ages of 20 and 24 (three male undergraduates and two female graduate students) participated in all experiments. All five listeners, denoted L1–L5, had participated in previous experiments in our laboratory but were selected solely on the basis of availability. They were paid for their participation and completed the experiments in five 2-h sessions (with breaks) over the course of 2 weeks.

2. General methods and procedures

The stimuli were generated at a 20-kHz sampling rate and low-pass filtered at 7.5 kHz. All masker bursts consisted of eight tones that were chosen randomly on a logarithmic frequency scale from the range 200–5000 Hz on every presentation, excluding the subregion 800–1250 Hz. The target was always contained within this protected subregion. Sounds were presented to listeners through matched TDH50 headphones while seated in individual sound attenuating rooms. Unless stated otherwise, the target was a 1000-Hz tone, the stimulus was presented monaurally to the right ear, and only one stimulus burst occurred in each interval of a trial. All experiments used a 2I, 2AFC two-down and one-up adaptive procedure with a fixed masker level of 60 dB per masker component (approximately 69 dB overall level) and an adaptive target level. Also, all experiments employed correct-answer, trial-by-trial, feedback.

The sessions began with several adaptive runs to estimate the unmasked target thresholds as well as to familiarize the listener with the targets. Next, masked thresholds were obtained alternating between the two conditions S and D after every two adaptive runs. A total of eight adaptive tracks, with a minimum of 50 trials and 9 reversals each, were obtained for every condition. Each adaptive track began with a step size of 4 dB that was changed to 2 dB after the third reversal. An even number of reversals, beginning with the fourth or fifth, were averaged to obtain one threshold estimate. To reduce learning effects, only the last 6 of the 8 threshold estimates were used in the final data analysis.

3. Specific stimuli

In the Duration experiment (Fig. 1, top), the duration of the eight-component masker was 300 ms (including 20 ms cosine-squared ramps for both onset and offset). In the S case, the target had exactly the same temporal characteristics as the masker. In the D case, the target began 100 ms later than the masker but retained the synchronous offset; thus its duration was only 200 ms (including its 20 ms cosine-squared ramps).

In the Sweep experiment (Fig. 1, bottom), the eight components of the masker had random starting frequencies (as in the above experiment), but instead of remaining constant they were rising frequency glides. The frequency of each

component increased by a factor of 1.49 over the 300-ms duration of the masker. In order to maintain the 5000-Hz upper limit on the frequencies present in the masker, the highest possible starting frequency of any masker component was 3356 Hz (5000 Hz/1.49). In the S case, the target was an upward glide from 820 to 1220 Hz. In the D case, it was a downward glide covering the same frequency range. In all cases, each component had a duration of 300 ms including 20 ms cosine squared onsets and offsets.

In the Spatial experiment (Fig. 2), the masker was the same as that used in the Duration experiment, except that it was presented diotically rather than monotically. The target tone was presented synchronously with the masker either diotically (the S case) or monotically (the D case).

In the MBS experiment (Fig. 3, top), the masker and the target consisted of eight contiguous 60-ms bursts with cosine-squared onset and offset ramps of 10 ms and a total duration of 480 ms. The frequency of each component of the masker was always held constant throughout the stimulus presentation (i.e., from burst to burst but not interval to interval). In the S condition, the frequency of the target was also held constant from burst to burst; in the D condition, however, the target frequency was randomly jittered from burst to burst over the range 820–1220 Hz.

In the MBD experiment (Fig. 3, bottom), the bursts have the same temporal characteristics as in the MBS case. However, in contrast to the MBS case, the frequencies of the masker were always randomized burst to burst in the range 200–5000 Hz (excluding the protected region). Thus, in this experiment the S condition employed a jittered-frequency target (in the range 820–1220 Hz) and the D condition employed a constant-frequency target.

IV. RESULTS

Possible learning effects were checked by examining the slopes of the threshold versus repetition functions. Averaged over the listeners and experiments, these slopes were relatively shallow (−0.38 dB/repetition for the S condition and −0.15 dB/repetition for the D condition). However, there was considerable variation in these slopes (the standard deviations were 2.2 and 1.8 dB/repetition, respectively). In general, these data are not adequate to study learning effects. Although for many listeners and many conditions, the learning observed over the last six repetitions (the measurements used in subsequent analyses) was relatively minor, one certainly cannot claim that asymptotic performance was approached. As pointed out previously, the issue of training in informational masking is an important one and will require substantial future work.

The average amount of masking for each experiment, displayed in bar graph form, is shown in Figs. 4 and 5. Whereas Fig. 4 shows the results averaged over listeners, Fig. 5 shows the results for the individual listeners (L1–L5). In both figures, the black bars show results for the S condition (target and masker similar) and the white bars for the D condition (target and masker dissimilar). In all cases, the amount of masking was obtained by subtracting each individual's unmasked target threshold from their masked threshold. Unmasked target thresholds ranged from −5 to 19 dB

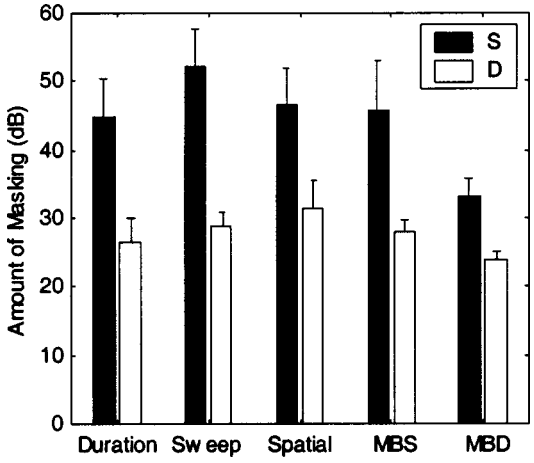


FIG. 4. Amount of masking (masked threshold minus unmasked threshold) for the five experiments and two target conditions S and D averaged over the five listeners. The results for the S condition are shown by black bars; the results for the D condition by white bars. The variation among listeners is indicated by the error bars, which show the standard error of the mean.

SPL for the various listeners and targets. The error bars in Fig. 4 give the standard error over the five listeners whereas the error bars in Fig. 5 give the standard deviation over the six final adaptive runs for each listener. (Standard error was chosen for Fig. 4 and standard deviation for Fig. 5 for visual display purposes. Standard deviations for Fig. 4 are obtained by multiplying the results shown by $\sqrt{5}$. Standard errors for Fig. 5 are obtained by dividing the results shown by $\sqrt{6}$.) On average, the standard deviations across the six repetitions are 7.1 dB for the S condition and 5.4 dB for the D condition. The most striking result seen in Fig. 4 is that in all five

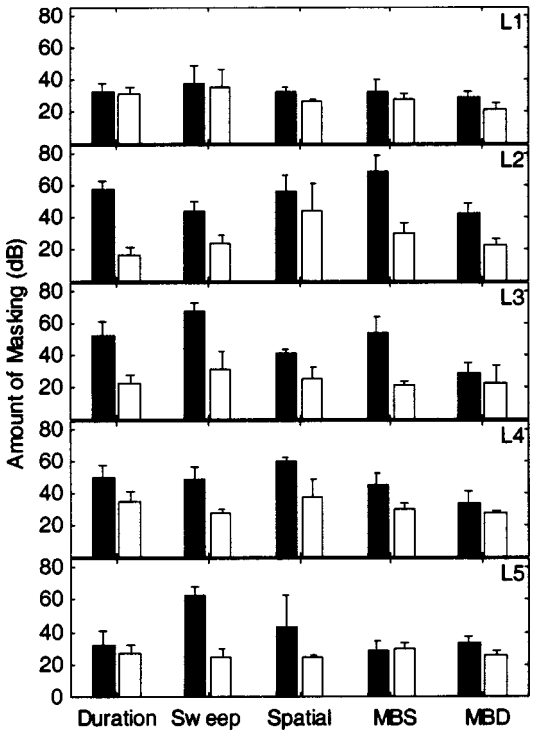


FIG. 5. The same as Fig. 4, except the results are plotted for each of the five listeners L1–L5 in separate panels. Here, the error bars give the standard deviation about the mean over the six adaptive runs used to estimate the threshold value.

TABLE I. This table shows the means (and standard deviations) for S, D, and S−D across experiments for each listener. The last column gives the means (and standard deviations) across listeners.

	L1	L2	L3	L4	L5	Mean (sd)
S	32.6(2.9)	53.9(10.8)	49.0(14.5)	47.3(9.5)	39.8(13.7)	44.5(8.4)
D	28.3(5.2)	27.6(10.6)	24.8(4.2)	31.5(4.6)	26.7(2.3)	27.8(2.5)
S−D	4.4(2.7)	26.3(12.9)	24.2(12.5)	15.8(6.4)	13.2(15.4)	16.8(8.8)

experiments there is considerably more masking for the S conditions than the D conditions. Given the specific target-masker parameters used and these specific five listeners, the most masking (average of approximately 52 dB) was obtained in the S condition for the Sweep experiment. The least masking for an S condition was obtained in the MBD experiment (approximately 33 dB). All of the D conditions produced less masking than their S counterparts and the order was somewhat different, with the Spatial experiment producing the most masking (approximately 32 dB) and the MBD experiment the least (an average of 24 dB). Despite these results and their statistical significance (discussed below), differences in the actual amount of masking obtained in the various experiments, the relative amount of masking across experiments, or the absolute size of the S−D difference will not be emphasized because of the dependence of these quantities on various arbitrary choices made in the design of the experiments (also discussed below).

Because listener differences are an important consideration (and we certainly do not assume that one can generalize the results obtained on five listeners to the population as a whole), a three-way ANOVA with listener as a factor was performed. The individual threshold estimates were used as the error term. All main effects and interactions were significant. Specifically, the three main factors of listener $[F(4,20)=9.97,\ p=0.0001]$, experiment $[F(4,20)=20.82,\ p<0.0001]$, and target-masker similarity $[F(1,5)=2951.3,\ p<0.0001]$ were all highly significant, indicating that each is an important determinant in the amount of masking produced. Perhaps more interesting are the interaction terms. The significant interaction of listener by experiment $[F(16,80)=10.27,\ p<0.0001]$ indicates that the difference across experiments depends on listener. The listener by target-masker similarity interaction $[F(4,20)=19.21,\ p<0.0001]$ indicates that the S vs D distinction also depends on listener, and the interaction of target-masker similarity and experiment $[F(4,20)=6.47,\ p=0.0017]$ indicates that the S vs D distinction also depends on experiment. The three-way interaction of listener by experiment by target-masker similarity $[F(16,80)=7.44,\ p<0.0001]$ indicates that the S vs D distinction depends on both listener and experiment. These points are elaborated below in connection with an examination of Figs. 4 and 5.

Looking first at Fig. 4, one sees that the average results across listeners clearly show the highly significant effect of target-masker similarity, as expected. The significant effect of experiment is also seen in that the experiments tend to produce different amounts of masking, with this difference appearing greater in the S condition than in the D condition. The interaction of experiment and target-masker similarity can be seen in that the effect of experiment is different for the S and D conditions. It is important to stress here, however, that a comparison of the amount of masking across experiments, or of the release from masking across experiments, calculated as S minus D (S−D), has limited meaning because these values could easily be changed by constructing the S conditions or the D conditions (or both conditions) differently. For example, of all the experiments performed, the Sweep experiment produces the most masking in the S condition and the greatest release from masking in going from the S condition to the D condition. However, the results for the Sweep experiment could obviously be radically changed merely by altering the slope of the sweep. Similarly, the results for the Duration experiment could be changed merely by altering the amount by which the target duration was shortened in the D condition. In general, no effort was made to equate (across experiments) either the effectiveness of the various maskers or the strength of the parameters used to produce the release from masking in the D conditions. Thus, in the remainder of this paper, comparisons of the amount of masking across experiments or the amount of release from masking across experiments will not be emphasized. Apart from the main result that in all experiments there is substantial release from masking in going from the S condition to the D condition, what is most interesting to explore in the data is how performance varied as a function of the listener, i.e., the results shown in Fig. 5.

Among the points to be noted when examining Fig. 5 are the following. First, as was implied in the average results, the D condition reduces the amount of masking relative to that obtained in the S condition in essentially all cases. Of the 25 comparisons between S and D shown in Fig. 5, 24 indicate a higher threshold for S than for D (the only exception is in the MBS experiment for L5). The mean values for S, D, and S−D (as well as the standard deviations) across experiments for each of the listeners (L1–L5) are shown in Table I along with the average and standard deviation of these mean values over the listeners. These results show quite clearly that the effect of target-masker similarity depends on listener (another of the significant two-way interactions in the ANOVA).

Second, not only is the amount of masking for D less than for S, but the intersubject variation in the amount of masking for D is less than for S. Although this result can be seen in Fig. 5, it is most clearly evident in the last column of Table I. The standard deviation across listeners for D is less than a third of the value for S (2.5 dB vs 8.4 dB).

Third, looking across both listeners and experiments in Fig. 5, one sees very large variations in the size of the S−D difference and in the amounts of masking for S and for D. The fact that the amount of masking in each experiment depends on the listener illustrates the significance of the two-way interaction of experiment and listener. Whereas, in some

TABLE II. This table shows the values of the quantity DP, which is a d' metric for measuring the difference between the results for the conditions S and D (taken from the values in Fig. 5), as well as the mean and standard deviation of DP across listener (last row) and experiment (last column). Bolded values are those differences that would be significant at the 0.05 level using a standard *t*-test given the Bonferroni correction (the critical DP value is 2.79, see text).

	Duration	Sweep	Spatial	MBS	MBD	Mean (sd)
L1	0.32	0.19	**2.89**	0.89	2.25	1.31 (1.2)
L2	**10.12**	**3.94**	0.96	**4.72**	**4.03**	4.76 (3.3)
L3	**4.64**	**4.82**	**3.05**	**5.10**	0.84	3.69 (1.8)
L4	2.27	**4.30**	**3.09**	**3.09**	1.22	2.79 (1.1)
L5	0.72	**8.04**	1.53	−0.33	1.92	2.38 (3.3)
Mean (sd)	3.61 (4.0)	4.26 (2.8)	2.31 (1.0)	2.69 (2.4)	2.05 (1.2)	

cases, the difference S−D is negligible, in other cases it is nearly 40 dB. The apparent dependence of this difference on both the experiment and the listener clearly reflects the statistical significance of the three-way interaction of these factors mentioned previously.

Fourth, it is evident in Fig. 5 that the variation among the error bars is extremely large (they vary by more than a factor of 10). In order to take these error bars into account when considering the difference S−D, a further analysis was performed. Specifically, in order to evaluate the S−D difference quantitatively, the d' metric DP, where

$$\mathrm{DP} = \frac{\overline{M}_S - \overline{M}_D}{\sqrt{(\sigma_S^2 + \sigma_D^2)/2}}, \tag{1}$$

was computed for each experiment and listener. In this expression, $\overline{M}_S$ and $\overline{M}_D$ denote the means and σ_S and σ_D denote the standard deviations for the S and D cases, respectively. The results of these computations are shown in Table II. The bold values in this table indicate the conditions that are statistically significant at the 0.05 level using a standard *t*-test and the Bonferroni correction. The critical value of *t* for df=10, $p=0.05$ (corrected to $p=0.05/25=0.002$) is 4.144, which corresponds to a critical DP of 2.79 (in these calculations DP is $t/1.58$). According to this conservative analysis, roughly half of the DP values are significant. Whereas L2–L4 have three or four significant differences, L1 and L5 have only one significant difference. Note also that whereas four of the five listeners show significant differences in the sweep experiment, only one listener has a significant difference in the MBD experiment. These same conclusions can be drawn by looking across the panels of Fig. 5 at each listener or down the columns of the figure at each experiment. This dependence of the target-masker similarity effect on listener, as well as the dependence of the target-masker similarity effect on experiment, was confirmed by the significance of both of these two-way interactions in the ANOVA results. Again, however, the variation of the results across experiments must be interpreted with extreme caution because of the lack of a natural metric for equating the various stimulus alterations used to transform the S condition to the D condition (the values of DP would undoubtedly change if different magnitudes for the parameter manipulations were used to create the D conditions). The statistically significant two-way interaction of experiment and target-masker similarity would not necessarily remain significant if each D manipulation were somehow constructed to produce reductions in masking that were more equal across experiments.

Fifth, and finally, the results shown in Fig. 5 indicate an intermediate level of consistency of individual subjects across experiments (normalized to the level of performance of the average subject across experiments in order to factor out the arbitrary aspect of the interexperiment comparisons). For example, on the side of consistency, note that whereas L1 has relatively small S−D differences (primarily because of relatively low values for S), L2–L4 tend to have relatively large differences (primarily because of relatively high values of S). On the side of inconsistency, however, note how the results for L5 are like those for L1 in the Duration and MBD experiments, but not in the Sweep and Spatial experiments. A much fuller and more quantitative analysis of listener consistency across experiments and of differences across listeners, a significant main effect in the ANOVA results, is presented in the Appendix. According to the results obtained in that analysis, the S condition is distinguishable from the D condition not only by the larger amount of masking and the larger amount of intersubject variation, but also by an increased tendency for knowledge of listener identity to improve predictive accuracy across experiments. More specifically, the rms deviation between the measured amounts of masking and the amounts predicted by a simple linear model that takes account of listener identity was calculated for both S and D conditions. To test the model, the obtained rms deviation was compared to the probability distribution of rms deviations that would occur by chance (i.e., by ignoring listener identity). For the S condition, the obtained rms deviation (or smaller) would occur only 3% of the time by chance, indicating a substantial degree of listener consistency across experiments. In contrast for the D condition an rms deviation less than or equal to that obtained by including listener identity would occur in 52% of the cases without this knowledge.

V. COMPARISONS WITH PREVIOUS DATA

As mentioned in Sec. II, some of these experiments are closely related to experiments reported by Kidd *et al.* (1994) and Neff (1995). Precise quantitative comparisons with these previous experiments cannot be made because, aside from differences in the set of listeners employed, there are substantial differences in the details of the experiments. For example, in the Neff study, unlike our study, a target cue was

always presented prior to each trial. Similarly, in the Kidd *et al.* multiple-burst experiments, unlike our multiple-burst experiments, the S condition was transformed into the D condition by altering the masker rather than the target. In addition, in neither study was the set of listeners held fixed across the experiments (thus preventing comparisons among studies of listener consistency across experiments). Nevertheless, to the extent that comparisons can be made across studies, the results appear relatively consistent. For example, using both a single-burst paradigm and a four-burst paradigm, for an eight-component masker in both a spatial experiment and a frequency-jitter experiment, Kidd *et al.* (1994) found large differences among listeners, with masking release in the range 0–40 dB (with an average of roughly 15 dB). Similarly, these same investigators found substantial release from informational masking (on the order of 20 dB) when the target was presented only during alternate bursts of the masker, despite the decrease in target energy in the alternate burst condition. In the Neff (1995) study, signal types, temporal factors, and spatial configuration were studied as a function of number of masker components. Even with the presence of a target cue immediately prior to each trial in all experiments (and trial-by-trial correct-answer feedback), substantial informational masking was obtained. Furthermore, and as expected, there was a substantial decrease in masking for most cases in which target-masker dissimilarity was introduced (again, differences among listeners were substantial). Relative to the baseline condition of a pure-tone target, AM (amplitude-modulated) targets, QFM (quasi-frequency-modulated) targets, and NBN (narrow-band noise) targets all showed decreased informational masking (although the QFM targets were least effective for this purpose). Most closely related to the experiments reported in this paper are the experiments with ten masker components in which the dissimilarity was created by shortening the signal duration or using different spatial configurations for the target and masker. The effect of making the target duration one-half the masker duration varied between 5 and 25 dB for the four listeners tested. Neff concluded that the duration manipulations were the most effective and most consistent at reducing the masking caused by masker frequency uncertainty. In that same study, the change in threshold in going from the monaural (or diotic) presentation for both target and masker to the case in which the target was presented interaurally out of phase varied over the range of 10–20 dB. As expected, a ''cross-ear'' condition in which the masker was in the ear opposite to the target produced the most release from masking although one listener required substantial practice before performance improved. In the spatial experiment by Kidd *et al.* (1994), for the case of 4 or 8 masker components, mean thresholds improved by 12–17 dB in going from the monotic to dichotic presentation (signal to one ear, masker to both ears in phase) for both the MBS and MBD (four-burst) conditions, and in going from the MBS to the MBD presentation for both the monotic and dichotic conditions.

For very rough comparisons with these previous data, the results of our experiments can be summarized as showing the following release-from-masking ranges and means (over the five listeners): a range of 6 to 22 dB and mean of 15 dB for Spatial; a range of 1 to 41 dB and mean of 18 dB for Duration; a range of −2 to 38 dB and mean of 18 dB for MBS; and a range 6 to 19 dB and mean of 9 dB for MBD. Because in both the experiments by Kidd *et al.* (1994) and Neff (1995) the listeners varied across experiments, the degree of listener consistency across experiments cannot be compared among the various sets of data.

VI. CONCLUDING REMARKS

The results reported in this paper, combined with those reported in the previous papers discussed above, clearly demonstrate that decreasing target-masker similarity (i.e., going from condition S to condition D) tends to reduce the masking effects of stimulus uncertainty. Also, as in previous experiments on informational masking, the intersubject variation is substantial. Furthermore, this variation appears much larger in the S condition than in the D condition. The amount of reduction (the threshold for S minus the threshold for D, S − D) depends both on the type of similarity change and on the listener. The results obtained in our experiments suggest that there is considerable structure in the matrix of thresholds for the ten different experimental conditions across the five different listeners. Specifically, in conditions where the amount of informational masking is reduced by decreasing target-masker similarity, individual differences in performance are relatively modest. In contrast, for conditions in which target and masker are similar, individual differences are large and the relative amounts of masking observed for a particular listener are moderately consistent across experiments that use different stimuli and methods of decreasing similarity. One cannot conclude, however, that individual differences are uniformly large in informational masking tasks and small in the reference tasks, or that a particular listener's vulnerability to informational masking is rigidly fixed across tasks. Such a conclusion would not only overstate the results obtained in this study, but other studies as well. For example, intersubject variability in informational masking tasks involving speech intelligibility seem somewhat reduced (Brungart, 2001; Arbogast *et al.*, 2002). Similarly, in some studies of informational masking, the intersubject differences in thresholds for the reference conditions (no uncertainty), as well as the informational masking conditions, appear quite large (e.g., Wright and Saberi, 1999; Durlach *et al.*, 2003b).

As indicated previously, it is difficult to draw conclusions about the relative potency in combating uncertainty of the different target-masker dissimilarity parameters introduced to convert the S condition to the D condition because of the arbitrary choice of the magnitudes of these parameters and the current lack of an independent metric to measure target-masker similarity. Despite this deficiency, all five of these particular experiments averaged across these particular five listeners produced substantial amounts of masking (between 33 and 52 dB) in the S conditions and large amounts of release from masking (9 to 23 dB) for the manipulations used to create the D conditions. Clearly, an important task for the future is to develop a target-masker similarity metric that can be applied to a wide variety of experimental situations.

A further issue that has not been addressed by the above experiments concerns the extent to which the observed releases from masking caused by the reductions of target-masker similarity in the various experiments would have occurred even if there had been no uncertainty in the masker. It has been implied implicitly by our use of the phrase ''combating uncertainty'' that if there were no uncertainty, there would be no nonenergetic masking for the decrease in target-masker similarity to combat. However, it is possible that even if the masker uncertainty had been totally eliminated, the decrease in target-masker similarity in going from condition S to condition D would have caused significant release from masking. In order to adequately explore this issue, it would have been necessary to measure thresholds for the S and D conditions in each experiment for all frozen exemplars of the random masker. To the extent that the results of this additional (massive) set of experiments showed a clear release from masking in going from condition S to condition D (and this release from masking were of sufficient magnitude to rule out explanations in terms of possible changes in energetic masking that might have occurred in some of the frozen cases in going from S to D), one would be forced either to define informational masking so as to include effects other than those associated with uncertainty OR to recognize that there exist types of nonenergetic masking other than informational masking. [More extended discussion of such definitional issues is available in Durlach *et al.* (2003a).]

It should also be noted that the data shown in Sec. IV of this paper cannot be compared to a quantitative theory of informational (or nonenergetic) masking because there is no such theory that now exists that takes quantitative account of both masker uncertainty and target-masker similarity. In order to develop such a theory, it will be necessary to define both uncertainty and similarity more adequately, determine improved methods for measuring these factors, and create a structure for properly integrating the effects of these factors.

It should further be noted that for such a theory to be truly successful, it will have to explain the very large individual differences observed as well as the effects of training (once these effects have been adequately documented empirically). Independent of whether the effects of training are generally large or small, and independent of the extent to which training tends to reduce the large individual differences observed, the study of training effects constitutes an essential step in the development of a serious theory. Although individual differences in susceptibility to informational masking are clearly of interest even if such differences can eventually be ''trained out,'' the way in which such differences should be modeled will obviously depend on the extent to which, and the manner in which, such ''training out'' can be achieved.

Finally, it should be noted that the stratagem in our research of focusing on simultaneous informational masking and temporarily ignoring sequential informational masking is not meant to imply that we believe that the latter area is unimportant or that an acceptable theory of informational masking can attend only to the simultaneous case. On the contrary, we believe that the results obtained on sequential masking (including results on individual differences and training effects) constitute a major building block in the search for an adequate theory. The area of sequential informational masking, and its relationship to simultaneous informational masking, will be considered in later papers.

ACKNOWLEDGMENTS

The authors thank Christine Carter and Sally Tressler for assistance with data collection and Dr. Robert Lutfi and Dr. Virginia Richards for useful discussions. In addition we are grateful to Dr. Marjorie Leek and two anonymous reviewers for their constructive suggestions. This work was supported by Grant Nos. DC00100 and DC04545 from NIH/NIDCD and Grant No. F49620-01-1-0005 from AFOSR.

APPENDIX: ANALYSIS OF LISTENER CONSISTENCY

One of the goals of the current set of experiments is to quantify the extent to which the large individual differences propagate across experiments. For example, if a particular listener has a very high threshold in the Duration experiment when target and masker have the same duration, does this listener also have a high a threshold in the Sweep experiment when both target and masker are swept in the same direction? How well does knowledge of individual performance in one task predict performance in another task, and is such individual knowledge more informative for S conditions (where individual differences are large and where informational masking is more important) than for D conditions? To what extent can individual listeners be characterized simply by determining their relative susceptibility to informational masking?

To begin to address these questions, we constructed a simple linear model of masking in which there is no interaction between the listener and the experiment (i.e., the effects of listener and experiment are completely separable[3]). We evaluated how well this model predicts the observed thresholds compared to predictions in which subject identity is ignored as well as predictions in which data were randomly permuted to destroy any listener consistency that might exist across experiments. For the S condition, the simple linear model that includes listener identity is shown to account for variability in the data beyond what one would expect by chance. However, for the D condition, the improvements in the model predictions that take into account listener identity are no more than would be expected by chance. These results, described below, suggest that listener differences are relatively consistent across different tasks when there is substantial informational masking, but not when there is little informational masking.

In the simple linear model, the amount of masking a particular listener exhibits in a particular experiment is assumed to be a sum of two factors: a factor specific to that experiment and a factor specific to that listener:

$$M(L,E)-\overline{M(L,E)}^{L,E}=(\overline{M(L,E)}^{L}-\overline{M(L,E)}^{L,E})+(\overline{M(L,E)}^{E}-\overline{M(L,E)}^{L,E}), \quad \text{(A1)}$$

TABLE III. This table shows the rms deviation (RMS), correlation (r), and correlation squared (r^2) for both the S condition and the D condition for Eq. (A2) and Eq. (A3).

	S condition			D condition		
Model	RMS	r	r^2	RMS	r	r^2
Eq. (A2)	7.8	0.78	0.61	4.8	0.57	0.33
Eq. (A3)	10.8	0.49	0.24	5.3	0.43	0.18

where L denotes the listener and can assume any of the values L1,L2,...,L5; E denotes the experiment and can assume any of the values Duration, Sweep, Spatial, MBS, or MBD; M(L,E) denotes the amount of masking (in dB) for listener L and experiment E (as shown in Fig. 5); $\overline{M(L,E)}^{L}$ denotes the average of M(L,E) over L (the group mean profile shown in Fig. 4); $\overline{M(L,E)}^{E}$ denotes the average of M(L,E) over E (as reported in Table I); and $\overline{M(L,E)}^{L,E}$ denotes the average of M(L,E) over both L and E (the grand mean of all the data as reported in the last column of Table I).

Note that by collecting and rearranging terms, Eq. (A1) can be rewritten simply as

$$M(L,E)=\overline{M(L,E)}^{L}+\overline{M(L,E)}^{E}-\overline{M(L,E)}^{L,E}. \quad (A2)$$

The relationship described by Eqs. (A1) and (A2) assumes that the results for listener L can be estimated by adding $\overline{M(L,E)}^{E}$ (a constant for each value of L) to the group-mean profile $\overline{M(L,E)}^{L}$, normalized by the overall group mean $\overline{M(L,E)}^{L,E}$. Note, furthermore, that equations (A1) and (A2) perfectly describe the data both when performance is the same for all listeners [because then $M(L,E)=\overline{M(L,E)}^{L}$ and $\overline{M(L,E)}^{E}=\overline{M(L,E)}^{L,E}$ for all L and E] and when performance is the same for all experiments [because then $M(L,E)=\overline{M(L,E)}^{E}$ and $\overline{M(L,E)}^{L}=\overline{M(L,E)}^{L,E}$ for all L and E].

In order to evaluate the extent to which Eq. (A2) represents the data for both the S and D conditions, the rms deviation between the predicted values of M(L,E) and the measured values of M(L,E) was computed (separately for S and D conditions). The results of this computation, included in Table III, show that the rms deviation for the S condition is 7.8 dB and the rms deviation for the D condition is 4.8 dB. If instead of using Eq. (A2) to estimate M(L,E), we used simply

$$M(L,E)=\overline{M(L,E)}^{L}, \quad (A3)$$

i.e., we ignored subject differences and just used the group-mean profile to estimate M(L,E), then the rms deviations (also shown in Table III) would have been 10.8 dB for the S condition and 5.3 dB for the D condition. Although in an absolute sense, the rms deviation between data and predictions is larger in the S condition than in the D condition, subject differences account for a larger percentage of the variation in the S condition [(10.8−7.8) dB out of 10.8 dB or 28%] than in the D condition [(5.3−4.8) dB out of 5.3 dB or 8.6%].

An alternative way to compare Eqs. (A2) and (A3) is to calculate the correlations between the predicted and actual results in each case and determine the percentage of variation in the data for which the model accounts. These calculations

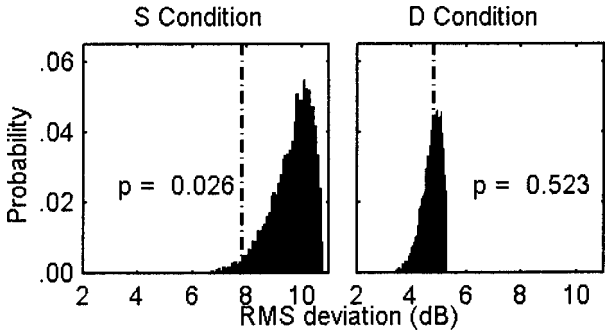


FIG. 6. Consistency of listeners across experiments. The figure shows the rms deviation between data and model for the both real listeners (dashed vertical line) and the pseudo listeners (probability density) for the S condition (left panel) and the D condition (right panel). See text for details.

(see Table III) show that 24% of the variance is accounted for in the S condition and 18% for the D condition when the mean alone is used [Eq. (A3)], but that these values increase to 61% in the S condition and 33% in the D condition when a listener-specific term is included in the predictions [Eq. (A2)]. Thus, incorporating knowledge of listener identity explains 37% more of the variance for the S condition, but only an additional 15% of the variance in the D condition (compared to using only knowledge of the experiment).

While these analyses suggest that knowledge of listener identity improves prediction accuracy, Eq. (A2) has more degrees of freedom than Eq. (A3); thus it is not a ''fair'' comparison. Even if data points for each experiment are randomly assigned to ''pseudo-listeners'' [rather than grouping the data by actual listeners in calculating $\overline{M(L,E)}^{E}$], the rms deviation will always decrease using the more-complex model [Eq. (A2)] compared to the experiment-only model [Eq. (A3)]. In order to obtain better insight into this issue, a boot-strapping method was used to determine the extent to which, for the data points we were fitting, the observed improvements in the model predictions is more likely to arise by chance than from the actual listener-specific characteristics. More specifically, in order to assess whether these improvements are better than expected by chance, we determined how often random permutations of the measured data lead to better predictions than the predictions based on grouping the data by listener. In other words, predictions using Eq. (A2) were compared to the results obtained when the correspondence between listeners and the measured values of M(L,E) were randomized. In this analysis, we (a) constructed results for randomized pseudo listeners by randomizing the correspondence between L and M(L,E) (subject only to the constraint that the experiment E was held fixed in the randomization); (b) calculated the rms deviation for each such randomization in the same manner as described previously; (c) performed 10 000 such randomizations and rms-deviation computations; and (d) used these results to estimate the probability density of the rms deviations for these randomized pseudo listeners.

The results of this analysis are shown in Fig. 6. For each of the conditions S and D, the figure shows both the rms deviation obtained with the real listeners (represented by the dashed vertical lines) and the estimated probability density

of the rms deviation for the pseudo listeners.[4] We also calculated the probability p of achieving an rms deviation with the pseudo listeners that is less than or equal to the actual rms deviation achieved with the real listeners (i.e., of achieving the actual rms deviation by chance). This quantity was calculated by estimating the area under the probability density functions to the left of the vertical line in Fig. 6. As shown, the probability of achieving this good a fit by chance is less than 3% in the S condition but is roughly 50% in the D condition. From this analysis, we conclude that compared to the D condition (in which the amount of informational masking is small), the results for the S condition (in which the amount of informational masking is large) show not only relatively large intersubject differences (see Table I), but also at least a modest degree of consistency across experiments. To what extent these results would continue to hold for other subjects and other experimental conditions is, of course, unknown.

Finally, it should be noted that although the relatively large rms deviations for the real listeners (7.8 dB for S and 4.8 dB for D) indicate that the model expressed by Eq. (A2) is far from perfect, even if the model were perfect these rms deviations would necessarily be substantially greater than zero because of the intrinsic noisiness of the M(L,E) measurements. In particular, note that if one averages the standard deviations of these measurements (displayed by the error bars in Fig. 5) across listeners and experiments, the results are 7.1 dB for condition S and 5.4 dB for condition D (corresponding to standard errors of 2.9 dB and 2.2 dB, respectively). With these numbers in mind, the rms deviations of 7.8 dB for S and 4.8 dB for D do not look so large.[5] In addition, more general models of listener consistency across experiments (nonlinear models, models incorporating parameters such as age, musical experience, etc.) may be more appropriate descriptions of how listener differences influence masking.[6] This simple linear representation was chosen because of its intuitive appeal and simplicity, not because we believe it represents the ''correct'' model of how listener differences propagate across experiments.

Overall, we interpret our results as confirming (1) the existence of large intersubject differences in susceptibility to informational masking and (2) substantial, but far from perfect, intrasubject consistency in susceptibility to informational masking across different types of informational masking experiments.

[1]It should be noted that we do not necessarily equate perceptual grouping (or fusion) with masking [see Bregman (1990) for a detailed discussion of this topic]. However, there is considerable evidence that when a target is perceptually grouped with an informational masker, the target is generally less detectable than when the stimuli are manipulated to promote segregation.

[2]The concept of masking due to perceptual similarity has been noted in the speech literature beginning at least with Egan *et al.* (1954) who distinguished between masking and ''confusion.'' Note also the close relationship between the terms ''perceptual masking'' [as used by Carhart *et al.* (1969)] and ''informational masking'' (as used by Freyman and colleagues and Brungart and colleagues in their recent speech-masking work).

[3]We realize that our data show some interaction between listener and experiment. This interaction is evident both when looking at Fig. 5 and when considering the results of the ANOVA. Nevertheless, it also appears that this interaction is sufficiently modest to warrant examination of listener consistency across experiments.

[4]Despite that fact that we refer to these randomizations as ''pseudo-listeners,'' it should be noted that they represent 10 000 variations of 25 values taken 5 at a time from the actual data obtained from only five listeners. This analysis is designed to examine whether the variations observed across the specific listeners used in this experiment are random or whether there are consistent differences among the results for these five particular subjects. This analysis is *not* equivalent to determining the variation that would be obtained if we actually observed 10 000 real listeners, but only considers whether the observed variation in the 25 data points we obtained is purely random or has some listener-specific structure.

[5]A rigorous evaluation of the model represented by Eq. (A2) would require modifying the computations underlying the results shown in Fig. 6 in such a way that the measurement noise is taken into account for both the real listeners and the pseudo listeners. We believe, however, that such a modification would add considerable complexity without providing much additional insight beyond that obtained with the simplified model used to derive the results shown in Fig. 6.

[6]A more general concept of listener consistency might require only that L and E be separable in the sense that functions H, f, and g exist such that M(L,E) can be well estimated by the equation $M(L,E)=H(\underline{f(L)},\underline{g(E)})$. In the model used in this paper [and ignoring the constant $\overline{M(L,E)^{L,E}}$], H merely adds f(L) and g(E), and f(L) and g(E) are chosen simply to equal $\overline{M(L,E)^{E}}$ and $\overline{M(L,E)^{L}}$, respectively. There is no guarantee, however, that this simple linear estimate is the best that could be found.

Arbogast, T. L., Mason, C. R., and Kidd, Jr., G. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Brungart, D. S. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' J. Acoust. Soc. Am. **110**, 2527–2538.

Carhart, R., Tillman, T., and Greetis, R. (**1969**). ''Perceptual masking of spondees by combinations of talkers,'' J. Acoust. Soc. Am. **45**, 694–703.

Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Colburn, H. S., Shinn-Cunningham, B. (**2003a**). ''Note on informational masking,'' J. Acoust. Soc. Am. (in press).

Durlach, N. I., Mason, C. R., Arbogast, T. L., and Kidd, Jr., G. (**2003b**). ''Informational masking: Psychometric functions for 'frozen' and 'mixed' maskers,'' in Abstracts of Twenty-sixth Annual Mid-Winter Meeting of the Association for Research in Otolaryngology, Vol. 26, p. 224.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). ''Some factors affecting multi-channel listening,'' J. Acoust. Soc. Am. **26**, 774–782.

Espinoza-Varas, B., and Watson, C. S. (**1989**). ''Perception of complex auditory patterns by humans,'' in *Comparative Psychology of Audition: Perceiving Complex Sounds*, edited by R. J. Dooling and S. H. Hulse (Erlbaum, Hillsdale, NJ).

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). ''The role of perceived spatial separation in the unmasking off speech,'' J. Acoust. Soc. Am. **106**, 3578–3588.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.

Kidd, Jr., G., Mason, C. R., and Arbogast, T. L. (**2002**). ''Similarity, uncertainty and masking in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, Jr., G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.

Leek, M., Brown, M. E., and Dorman, M. F. (**1991**). ''Informational masking and auditory attention,'' Percept. Psychophys. **50**, 205–214.

Lutfi, R. A. (**1990**). ''How much masking is informational masking,'' J. Acoust. Soc. Am. **88**, 2607–2610.

Lutfi, R. A. (**1993**). ''A model of auditory pattern analysis based on component-relative-entropy,'' J. Acoust. Soc. Am. **94**, 748–758.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). ''One factor underlies individual differences in auditory informational masking within and across age groups,'' Percept. Psychophys. **65**, 396–406.

Neff, D. L. (**1995**). ''Signal properties that reduce masking by simultaneous, random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.
Neff, D. L., and Callaghan, B. P. (**1987**). *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 37–46.
Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.
Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.
Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). ''Informational masking for multicomponent maskers with spectral gaps,'' J. Acoust. Soc. Am. **94**, 3112–3126.
Oh, E. L., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**, 3489–3499.
Oh, E. L., and Lutfi, R. A. (**2000**). ''Effect of masker harmonicity on informational masking,'' J. Acoust. Soc. Am. **108**, 706–709.
Oxenham, A. J., Fligor, B., Mason, C. R., and Kidd, Jr., G. (**2003**). ''Informational masking and musical training,'' (in review).
Richards, V. M., Tang, Z., and Kidd, Jr., G. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.
Turvey, M. T. (**1973**). ''On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli,'' Psychol. Rev. **80**, 1–52.
Watson, C. S. (**1987**). ''Uncertainty, informational masking and the capacity of immediateauditory memory,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.
Watson, C. S., and Kelly, W. J. (**1981**). ''The role of stimulus uncertainty in the discriminability of auditory patterns,'' in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard, Jr. (Erlbaum, Hillsdale, NJ), pp. 37–59.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1186.
Wright, B. A., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.

Effectiveness of narrow-band versus tonal off-frequency maskers

Sophie Savel and Sid P. Bacon
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 27 November 2002; revised 22 April 2003; accepted 25 April 2003)

The present study was a follow-up to a pilot study in which it was found that a 500-Hz-wide narrow-band noise (NBN) masker produced more masking than a tonal (T) masker for signal frequencies both above and below the masker frequency. The aim of the present study was to determine to what extent these results were influenced by an interaction of the relatively rapid temporal envelope fluctuations of the NBN and the short (10-ms) duration of the signal. In the first experiment, the masking produced by a regular NBN, a low-noise noise (LNN), and a T was compared. The LNN produced less masking than the NBN, and about as much as the T, suggesting that the inherent amplitude fluctuations in the NBN were largely responsible for the greater masking produced by that masker. In the second experiment, the masking produced by a regular NBN was compared with that by a T for a signal duration of 10 or 200 ms. The difference in masking between the two maskers was reduced or eliminated when the signal duration was 200 ms, because the threshold in the presence of the NBN masker decreased more with increasing signal duration. This could reflect a decreased ''confusion'' between the signal and the inherent fluctuations of the NBN masker. [DOI: 10.1121/1.1582442]

PACS numbers: 43.66.Dc, 43.66.Mk [MRL]

I. INTRODUCTION

The amount of masking produced by a bandlimited masker can depend significantly upon the bandwidth of that masker and the frequency relation between the signal and masker. For example, when the signal frequency (f_s) is about $\frac{1}{2}$ octave or more above the masker (center) frequency (f_m), narrow-band noise (NBN) maskers can produce considerably less masking than tonal (T) maskers (Egan and Hake, 1950; Buus, 1985; Mott and Feth, 1986; van der Heijden and Kohlrausch, 1995). This difference is likely related to differences in the temporal envelope of the two maskers (Buus, 1985; Mott and Feth, 1986; van der Heijden and Kohlrausch, 1995), which is flat for T maskers but fluctuating for NBN maskers. Subjects can presumably ''listen in the dips'' of the fluctuating masker to improve signal detection. For wider noise bandwidths, and hence faster fluctuations, the difference in masking effectiveness largely disappears (Buus, 1985). When f_s is less than about $\frac{1}{2}$ octave or so above f_m, on the other hand, NBN maskers may produce more masking (Egan and Hake, 1950), although fewer studies have evaluated this relationship, and the difference in masking may depend upon whether or not combination products are masked (Mott and Feth, 1986; Moore *et al.*, 1998).

When f_s is below f_m, NBN maskers often produce more masking than T maskers (Egan and Hake, 1950; Mott and Feth, 1986). This difference, however, may also depend upon masker bandwidth. Glasberg and Moore (1994) found that the two masker types produced about the same amount of masking when the noise masker was only 16 Hz wide, but that the NBN masker produced more masking when its bandwidth was considerably wider (about a critical band or more wide). They argued that the greater masking by the wider masker was due to combination products generated by the interaction of noise components within the NBN.

We recently collected data from a pilot study designed to measure temporal effects in simultaneous masking that showed that a NBN masker produced 8–10 dB more masking than a T masker, regardless of whether the signal was lower or higher in frequency than the masker. The masker–signal frequency separation was only about 1 ERB (equivalent rectangular bandwidth; Glasberg and Moore, 1990), corresponding to much less than $\frac{1}{4}$ oct, and thus these results may not be inconsistent with results in the literature, although as noted above it is somewhat unclear what to expect when f_s is just somewhat higher in frequency than f_m (Mott and Feth, 1986). In that pilot experiment, the width of the NBN masker was 500 Hz and the signal duration was 10 ms. Thus, the noise was wider and the signal was shorter than in most studies comparing the masking effectiveness of NBN and T maskers. The relatively rapid fluctuations of the wider noise masker may have interacted with the short signal to yield especially large amounts of masking. The goal of the present study was to evaluate this possibility. This was accomplished in two ways. The first way was by comparing the masking produced by a regular NBN, a low-noise noise (LNN), and a T. With LNN (Pumplin, 1985; Kohlrausch *et al.*, 1997), fluctuations in the temporal envelope are minimized. The second way was by comparing the masking produced by a regular NBN with that produced by a T for a signal duration of either 10 or 200 ms.

II. METHOD

A. Apparatus and stimuli

The signal and masker were digitally generated and produced at a 50-kHz sampling rate using a digital array pro-

 0001-4966/2003/114(1)/380/6/$19.00

cessing card (TDT AP2) and digital-to-analog converter, or DAC (TDT DD1). The output of the DAC was low-pass filtered at 8 kHz (TDT FT6).

The signal was a 4-kHz tone; its duration was 10 or 200 ms. The masker was either an NBN with a width of 500 Hz and a closer cutoff frequency that was 1.038 ERBs from f_s, a LNN whose cutoff frequencies were identical to those of masker NBN, or a T whose frequency was equal to the closer cutoff frequency of the noise maskers. The masker was 400 ms in duration. It was located either below or above the signal in frequency. For $f_m < f_s$, the lower and upper cutoff frequencies for NBN and LNN were 3.053 and 3.553 kHz, respectively, and the frequency of the T was 3.553 kHz. For $f_m > f_s$, the lower and upper cutoff frequencies for NBN and LNN were 4.5 and 5.0 kHz, respectively, and the frequency of the T was 4.5 kHz. The NBN was generated in the frequency domain and then transformed into the time domain via an inverse Fourier transform. A new sample of the NBN masker was generated on each trial. The LNN was generated according to the algorithm described by Kohlrausch *et al.* (1997). Only one version of the LNN masker was generated for each of the two center frequencies (i.e., for a given center frequency, the LNN was ''frozen''). The waveforms of the NBN and LNN maskers were analyzed and their differences were identical to those described in earlier studies (e.g., Kohlrausch *et al.*, 1997, Fig. 2).

The signal and masker were each gated with 5-ms $\cos^2$ rise/fall times. The stimulus durations (10 or 200 ms for the signal, 400 ms for the masker) included these rise/fall times. The signal was presented at the temporal center of the masker. The level of the signal was varied adaptively via the array processor. For the most part, the overall level of the masker was 62 or 82 dB SPL.

When $f_m < f_s$, a low-pass noise was presented continuously to avoid the detection of combination tones or bands. The noise (TDT WG1) was low-pass filtered (TDT PF1) at a cutoff frequency equal to the lower edge of the masker band for maskers NBN and LNN, and equal to the CDT ($2f_m - f_s$) for masker T. The spectrum level of the noise was 10 (62-dB masker) or 30 dB SPL (82-dB masker). The masker and low-pass noise were attenuated separately (TDT PA4), added to the signal (TDT SM3), and routed via a headphone buffer (TDT HB6) to one earpiece of Sennheiser HD250 headphones.

B. Procedure

Subjects were tested in a single-walled, sound-attenuating booth located within a sound-insulated room. Thresholds were measured using an adaptive, two-interval, forced-choice procedure that estimates the 79.4%-correct point on the psychometric function (Levitt, 1971). Lights were used to indicate when the signal might occur and to provide correct-answer feedback. The two observation intervals were separated by 500 ms. A run started with the signal level 10–15 dB above the estimated threshold; it was decreased following three correct responses and increased following one incorrect response. The step size was initially 5 dB, but was decreased to 2 dB following the second reversal. Each run consisted of 12 reversals; the threshold estimate for that run was the mean level at the last 10 reversals. A run was discarded if the standard deviation of the threshold estimate was greater than 5 dB; this rarely occurred. Each threshold reported is the mean of at least three runs obtained over 3 separate days. If the standard deviation of this mean exceeded 3 dB, an additional estimate was obtained and included in the mean. Over 96% (139/144) of the means presented here had standard deviations less than 3 dB.

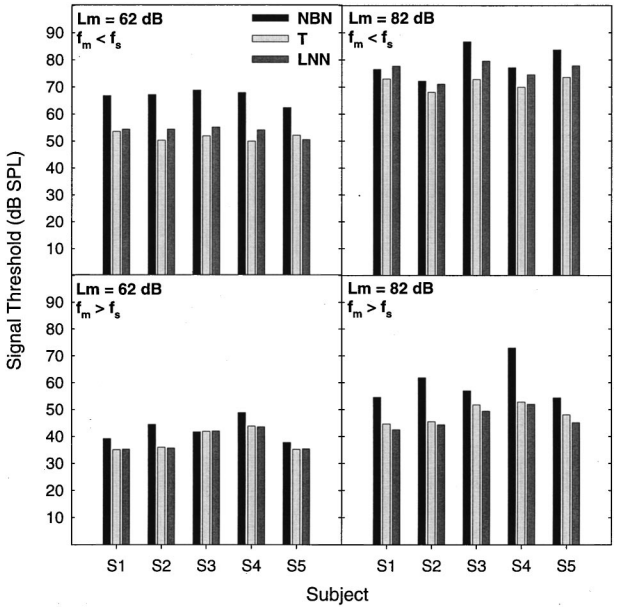


FIG. 1. Individual masked thresholds obtained with a 500-Hz-wide narrow-band noise masker (NBN: black bars), a tonal masker whose frequency was equal to the closer cutoff frequency of the NBN masker (T: light-gray bars), and a low-noise noise masker whose cutoff frequencies matched those of the NBN masker (LNN: dark-gray bars). Results for $f_m < f_s$ are presented on the first row; those for $f_m > f_s$ are presented on the second row. The left and right columns present data obtained with a 62-dB and an 82-dB masker, respectively. Signal duration was 10 ms.

C. Subjects

Five individuals participated. They ranged in age from 21–29 years, and had thresholds of 15 dB HL or lower (ANSI, 1996) for octave frequencies from 0.25 to 8.0 kHz. All subjects except S2 had prior experience in simultaneous-masking experiments.

III. RESULTS

A. Effect of LNN on the noise-tone difference

Figure 1 presents individual masked thresholds for the 10-ms signal obtained with maskers NBN, T, and LNN, respectively. The results for $f_m < f_s$ are in the two upper panels, whereas those for $f_m > f_s$ are in the two lower panels. The left panels present data obtained with the 62-dB masker, whereas the right panels present those obtained with the 82-dB masker. The thresholds obtained with masker NBN were generally higher than those obtained with masker T, although the magnitude of the difference depended somewhat upon masker level and masker–signal frequency relation. Averaged across subjects and conditions, the NBN masker produced 10 dB more masking than the T masker.

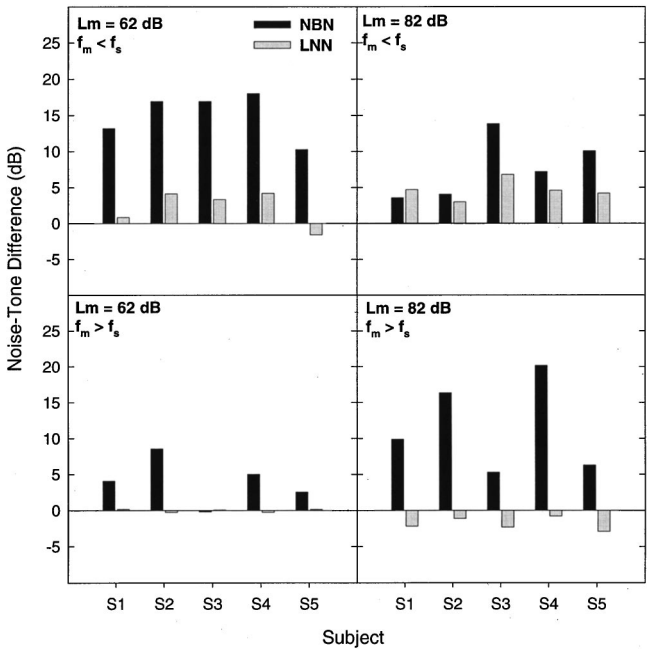


FIG. 2. The results from Fig. 1 plotted as the difference between the masked threshold obtained with the noise masker (NBN: black bars; LNN: gray bars) and that obtained with the T masker.

This difference in masking decreased in all conditions with the LNN. Indeed, in most conditions the difference nearly disappeared when the NBN was replaced with the LNN. The exception was in the condition where the 82-dB masker was lower in frequency than the signal (top right panel), in which case the LNN still produced somewhat more masking than the T. Averaged across subjects and conditions, the LNN masker produced only 1 dB more masking than the T masker[1].

Figure 2 presents the same set of data, but expressed in terms of the difference in masking between the noise (NBN or LNN) and T maskers. On these coordinates, it is easy to see that the noise-tone difference is much smaller for the LNN than for the NBN. For conditions where $f_m < f_s$, the noise-tone difference with LNN was only about 3 dB, whereas for conditions where $f_m > f_s$, this difference was essentially 0 dB (i.e., the LNN and T maskers were nearly equally effective).

For the NBN masker, the noise-tone difference depended upon masker level. For $f_m < f_s$, it decreased with level (except for S5, for whom the difference was the same at both levels). For $f_m > f_s$, on the other hand, it increased with level. To examine these effects of level further, we measured growth-of-masking functions with NBN and T maskers for two subjects (S1 and S5). Figure 3 shows the results for $f_m < f_s$. The overall masker level ranged from 42 to 92 dB in 10-dB steps. The same low-pass noise that was used in the main experiment to mask combination products was used here. It was presented continuously in the test ear at a spectrum level that was 52 dB lower than the masker level. The arrows in the figure indicate the masker levels used in the main experiment (these conditions were rerun for these functions). Each function in Fig. 3 was fitted with two lines chosen to minimize the squared deviations (points representing

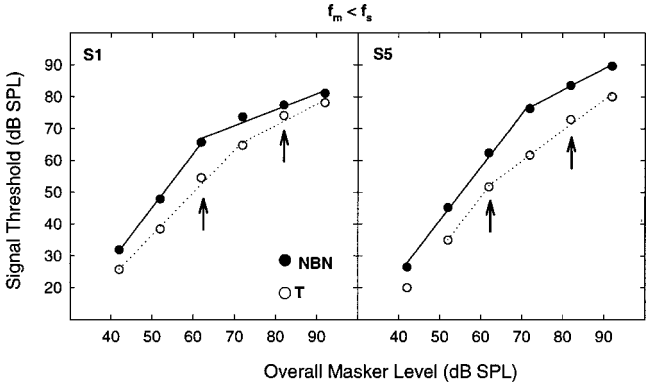


FIG. 3. Growth-of-masking functions for $f_m < f_s$ for masker NBN (filled circles) and masker T (unfilled circles). The signal duration was 10 ms. The results for subjects S1 and S5 are shown in separate panels. The arrows indicate the masker levels that were used for the data in Figs. 1 and 2.

less than 5 dB of masking were excluded from the fits). The slope of the initial line (with masker NBN: 1.69 for S1, 1.67 for S5; with masker T: 1.33 for S1, 1.68 for S5) was steeper than that of the second line (with masker NBN: 0.49 for S1, 0.67 for S5; with masker T: 0.67 for S1, 0.96 for S5). This decrease in slope at high levels has been observed previously (e.g., Schöne, 1977; Gregan *et al.*, 1998; Bacon *et al.*, 1999). For S1, the decrease occurred at about the same signal level (65 dB SPL) for both maskers, which corresponded to a higher masker level for the T masker than for the NBN masker. Consequently, the difference between the two masking functions (the noise-tone difference) decreased at high masker levels. For S5, the two functions did not change slope at the same signal (or masker) level, and for this subject the noise-tone difference did not change much with changes in masker level. Although masking functions were not obtained for S2, S3, and S4, their data in Figs. 1 and 2 suggest that their functions would be more like those of S1 than S5.

Figure 4 shows the growth-of-masking functions for S1 and S5 for $f_m > f_s$. Each function was fitted with a single straight line. For both maskers, the slope was less than unity, consistent with previous studies on simultaneous masking using maskers higher in frequency than the signal (e.g., Wegel and Lane, 1924; Egan and Hake, 1950). The slope of the function obtained with masker NBN (0.71 for S1, 0.75 for S5) was steeper than that obtained with masker T (0.54 for

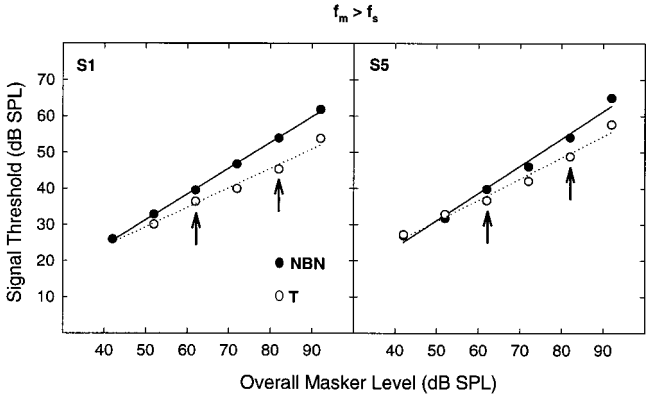


FIG. 4. Same as Fig. 3, but for $f_m > f_s$.

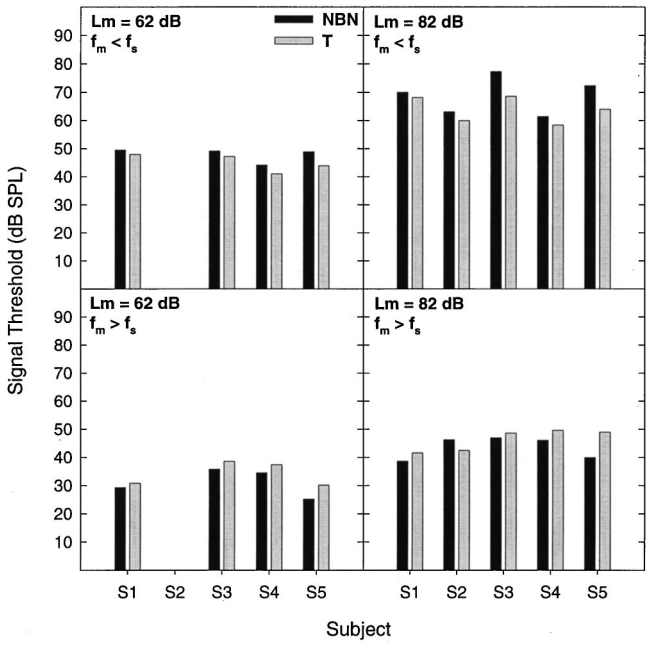


FIG. 5. Individual masked thresholds obtained with maskers NBN (black bars) and T (gray bars) for a 200-ms signal.

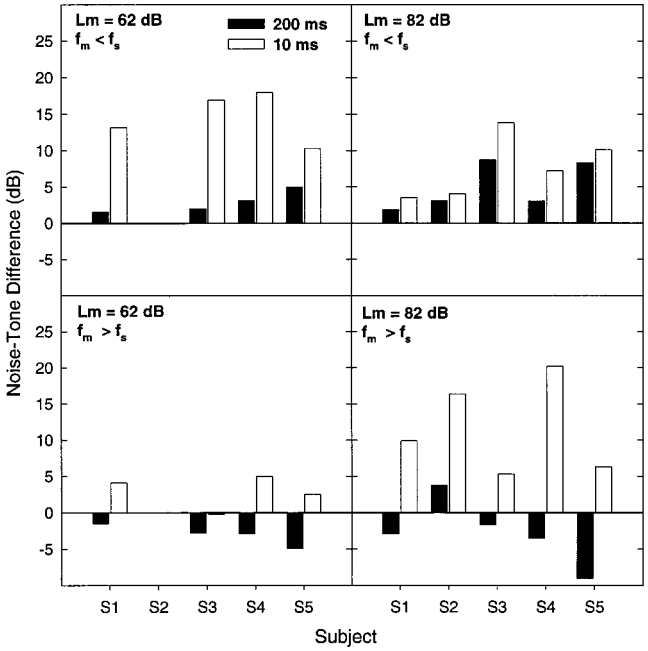


FIG. 6. The results from Fig. 5 plotted as the difference between the masked thresholds obtained with the NBN and T maskers (black bars). The results obtained with a 10-ms signal (white bars) are replotted from Fig. 2 for comparison.

S1, 0.59 for S5). This difference in slope is in good agreement with results from Glasberg and Moore (1994), and it can account for the increase in noise-tone difference with increasing masker level observed in Fig. 2 for $f_m > f_s$.

B. Effect of signal duration on the noise-tone difference

Figure 5 presents masked thresholds for the 200-ms signal obtained with maskers NBN and T. Subject S2 was not available to be tested with the 62-dB masker. With a few exceptions, the signal levels at threshold obtained in the presence of masker NBN were about the same as those obtained in the presence of masker T, regardless of masker level and masker–signal relation.

Figure 6 shows the same set of data expressed in terms of the noise-tone difference. The results obtained with the 200-ms signal (black bars) are compared with those obtained with the 10-ms signal (white bars; replotted from Fig. 2). For $f_m < f_s$, the noise-tone difference averaged across subjects and masker level with the long signal was 4 dB. For the short signal, the average difference was 11 dB. For $f_m > f_s$, most subjects showed a slightly negative noise-tone difference with the long signal at both masker levels. On average (across subjects and masker level), the difference was -3 dB. For the short signal, the average noise-tone difference was 8 dB.

IV. DISCUSSION

A. Fluctuating-envelope masker versus flat-envelope maskers

The goal of the present study was to evaluate the extent to which rapid envelope fluctuations of a relatively wide NBN masker interact with a short signal to yield relatively large amounts of masking. In one set of conditions, we compared the masking produced by a regular NBN, an LNN, and a T. The purpose was to determine whether the relatively rapid, inherent fluctuations of the NBN might be at least partly responsible for our previous finding that an NBN masker can produce more masking than a T masker. We found that a 500-Hz-wide LNN produces less masking than a regular NBN with identical cutoff frequencies and overall level. Moreover, the masking produced by the LNN was about the same as that produced by the T, both when the masker was lower and when it was higher in frequency than the signal. This suggests that the greater masking produced by the regular NBN masker was primarily related to the inherent amplitude fluctuations of that masker.

Another possible explanation for the difference in masking observed between the NBN (fluctuating-envelope) masker and either the T or LNN (flat-envelope) masker is related to informational masking. In particular, the greater uncertainty associated with the NBN masker (which varied from trial to trial) may have caused that masker to be a better masker. Although we cannot rule out the possibility that masker uncertainty contributed at least somewhat to our results[2], previous results on the effects of signal duration on informational masking (Neff, 1995) suggest that it is highly unlikely that our major findings can be attributed to masker uncertainty. Neff showed that the amount of informational masking *decreased* as the duration of a signal temporally centered within a 200-ms masker decreased from 100 to 10 ms. On the contrary, we found that the difference in masking produced by the 400-ms NBN and T maskers *increased* as the duration of the temporally centered signal decreased from 200 to 10 ms. In other words, the effect of signal duration observed here is exactly opposite to what would be expected if the greater masking produced by the NBN masker were due to informational masking.

B. Effect of signal duration

In a second set of conditions, we compared the masking produced by a regular NBN with that produced by a T for a signal duration of either 10 or 200 ms. This was done to determine whether the greater masking effectiveness of the NBN masker that we observed in our pilot data might be due to our use of a short signal. Before we discuss why we believe the NBN was more effective than the T in masking short signals, we first discuss how our results with a long signal compare with those in the literature.

1. Comparison with the literature

In the present study, for $f_m < f_s$, the regular NBN masker produced slightly more masking than the T masker when the signal duration was 200 ms (see Fig. 6). These results are consistent with those of Mott and Feth (1986), despite the differences in masker bandwidth (we used 500 Hz, they used 50 Hz). For $f_m > f_s$, the results presented here showed that a T masker produced more masking of a 200-ms signal than did an NBN masker. This is in contrast to the results of earlier studies, where T maskers produced less masking than NBN maskers when the masker was higher in frequency than the signal (e.g., Mott and Feth, 1986; Glasberg and Moore, 1994). This difference may be related to how the T and NBN maskers were ''equated'' in frequency. In the previous studies, the frequency of the T masker was equal to the center frequency of the NBN. In the present study, it was equal to the closer (in this case, lower) cutoff frequency of the NBN masker. Consequently, for a given nominal masker–signal distance, the T masker in the present study was located closer to the signal than it was in the previous studies. This may explain why the T masker was relatively more effective in the present study, and hence why the noise-tone difference was consistently negative.

2. Why signal duration is more important with the NBN masker

We can think of two reasons why the difference in masking between the NBN and T maskers was generally larger with the short signal than with the long signal. Both pertain to why the masked threshold decreases with increasing signal duration more in the presence of the NBN masker than in the presence of the T masker (compare thresholds in Fig. 1 with those in Fig. 5). The first is related to the fact that the short signal rarely would be located in a masker minimum, whereas the longer signal would persist through several minima, thus providing numerous opportunities for the subject to take advantage of the momentary increases in signal-to-masker ratio that exist in those minima. Although plausible, this explanation is inconsistent with the finding that the NBN masker produced more masking of the long signal than the T masker for conditions where $f_m < f_s$ (top row of Fig. 6). Furthermore, the rapid fluctuations inherent in the NBN masker may not have provided sufficient opportunities to ''listen in the dips'' anyway (Buus, 1985). The second is related to the possibility that subjects had difficulty distinguishing the brief signal from the inherent fluctuations of the NBN masker. This ''confusion'' argument is along the same lines as those used by others to explain, for example, why gap-detection thresholds are especially large with NBN markers (e.g., Moore and Glasberg, 1988), or why forward-masked thresholds can be especially high under certain stimulus conditions (Weber and Moore, 1981; Neff, 1985, 1986). Consistent with this argument is the finding that the average time between envelope minima of a 500-Hz NBN is similar in duration (6.2 ms)[3] to our short signal (10 ms).

V. SUMMARY

(1) In simultaneous masking, a 500-Hz-wide NBN masker produces more masking of a short signal than a T masker of the same overall level whose frequency is equal to the closer cutoff frequency of the NBN. This was observed for maskers on both the low- and high-frequency sides of the signal.

(2) The noise-tone difference in masking effectiveness is reduced or eliminated when the NBN is replaced with a LNN with the same cutoff frequencies and overall level. This indicates that the greater masking by the NBN is related to its inherent envelope fluctuations.

(3) The difference in masking effectiveness between an NBN masker and a T masker is largely reduced ($f_m < f_s$) or even reversed ($f_m > f_s$) with a long signal.

(4) The relatively large effect of signal duration in the presence of the NBN masker may be related to the fact that the longer signal is less likely to be ''confused'' with the inherent fluctuations in the masker envelope.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute of Deafness and Other Communication Disorders (NIDCD Grant No. DC01376). We thank Marjorie Leek, Søren Buus, and an anonymous reviewer for their helpful comments on an earlier version of this manuscript.

[1]As only one sample of the LNN masker was used for each of the two masker–signal relationships, it could be argued that the results obtained with the LNN are not representative, but rather are specific to the particular samples used. To address this issue, additional threshold estimates for subject S1 with $f_m > f_s$ at a masker level of 82 dB SPL were obtained with three new samples of the LNN masker. Those thresholds (42.0, 41.4, and 41.8 dB SPL) were nearly identical to that shown in Fig. 1 (43.4 dB SPL). This strongly suggests that our results with the LNN masker were not an artifact of the particular sample of LNN that we used.

[2]Additional data collected from S1 (82-dB masker, $f_m > f_s$) with a 200-ms signal suggest that informational masking may contribute somewhat to the greater masking effectiveness of the NBN masker. The threshold with the NBN masker (38.6 dB) was slightly higher than that obtained with the LNN masker (36.2 dB).

[3]Envelope statistics were obtained on an NBN as follows: a 1-s sample of the noise was generated with cutoff frequencies identical to those used in the experiment. The envelope of the noise was extracted via half-wave rectification and low-pass filtering at a cutoff frequency of 250 Hz (bandwidth/2). Approximately 160 envelope minima were identified. The most frequent time between successive minima was 6.2 ms.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for Audiometers'' (American National Standards Institute, New York).

Bacon, S. P., Boden, L. N., Lee, J., and Repovsch, J. L. (**1999**). ''Growth of simultaneous masking for $f_m < f_s$: Effects of overall frequency and level,'' J. Acoust. Soc. Am. **106**, 341–350.

Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Egan, J. P., and Hake, H. W. (**1950**). ''On the masking pattern of a simple auditory stimulus,'' J. Acoust. Soc. Am. **22**, 622–630.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Glasberg, B. R., and Moore, B. C. J. (**1994**). ''Growth-of-masking functions for several types of maskers,'' J. Acoust. Soc. Am. **96**, 134–144.
Gregan, M. J., Bacon, S. P., and Lee, J. (**1998**). ''Masking by sinusoidally amplitude-modulated tonal maskers,'' J. Acoust. Soc. Am. **103**, 1012–1021.
van der Heijden, M., and Kohlrausch, A. (**1995**). ''The role of envelope fluctuations in spectral masking,'' J. Acoust. Soc. Am. **97**, 1800–1807.
Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., and Oxenham, A. J. (**1997**). ''Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations,'' Acustica **83**, 659–669.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Moore, B. C. J., Alcantara, J. I., and Dau, T. (**1998**). ''Masking patterns for sinusoidal and narrow-band noise maskers,'' J. Acoust. Soc. Am. **104**, 1023–1038.
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.
Mott, J. B., and Feth, L. L. (**1986**). ''Effects of the temporal properties of a masker upon simultaneous masking patterns,'' in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 381–386.
Neff, D. L. (**1985**). ''Stimulus parameters governing confusion effects in forward masking,'' J. Acoust. Soc. Am. **78**, 1966–1976.
Neff, D. L. (**1986**). ''Confusion effects with sinusoidal and narrow-band noise forward maskers,'' J. Acoust. Soc. Am. **79**, 1519–1529.
Neff, D. L. (**1995**). ''Signal properties that reduce masking by simultaneous, random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.
Pumplin, J. (**1985**). ''Low-noise noise,'' J. Acoust. Soc. Am. **78**, 100–104.
Schöne, P. (**1977**). ''Nichtlinearintäten im mithörschwellen-tonheitsmuster von sinustönen,'' Acustica **37**, 37–44.
Weber, D. L., and Moore, B. C. J. (**1981**). ''Forward masking by sinusoidal and noise maskers,'' J. Acoust. Soc. Am. **69**, 1402–1409.
Wegel, R. L., and Lane, C. E. (**1924**). ''The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear,'' Phys. Rev. **23**, 266–285.

The effects of the amplitude distribution of equal energy exposures on noise-induced hearing loss: The kurtosis metric

Roger P. Hamernik,[a] Wei Qiu, and Bob Davis
Auditory Research Laboratory, State University of New York, 107 Beaumont Hall, Plattsburgh, New York 12901

(Received 29 July 2002; revised 24 January 2003; accepted 25 April 2003)

Seventeen groups of chinchillas with 11 to 16 animals/group ($\Sigma N = 207$) were exposed for 5 days to either a Gaussian (G) noise or 1 of 16 different non-Gaussian (non-G) noises at 100 dB(A) SPL. All exposures had the same total energy and approximately the same flat spectrum but their statistical properties were varied to yield a series of exposure conditions that varied across a continuum from G through various non-G conditions to pure impact noise exposures. The non-G character of the noise was produced by inserting high level transients (impacts or noise bursts) into the otherwise G noise. The peak SPL of the transients, their bandwidth, and the intertransient intervals were varied, as was the rms level of the G noise. The statistical metric, kurtosis (β), computed on the unfiltered noise $\beta(t)$, was varied $3 \leqslant \beta(t) \leqslant 105$. Brainstem auditory evoked responses were used to estimate hearing thresholds and surface preparation histology was used to determine sensory cell loss. Trauma, as measured by asymptotic and permanent threshold shifts (ATS, PTS) and by sensory cell loss, was greater for all of the non-G exposure conditions. Permanent effects of the exposures increased as $\beta(t)$ increased and reached an asymptote at $\beta(t) \sim 40$. For $\beta(t) > 40$ varying the interval or peak histograms did not alter the level of trauma, suggesting that, in the chinchilla model, for $\beta(t) > 40$ an energy metric may be effective in evaluating the potential of non-G noise environments to produce hearing loss. Reducing the probability of a transient occurring could reduce the permanent effects of the non-G exposures. These results lend support to those standards documents that use an energy metric for gauging the hazard of exposure but only after applying a ''correction factor'' when high level transients are present. Computing β on the filtered noise signal $[\beta(f)]$ provides a frequency specific metric for the non-G noises that is correlated with the additional frequency specific outer hair cell loss produced by the non-G noise. The data from the abundant and varied exposure conditions show that the kurtosis of the amplitude distribution of a noise environment is an important variable in determining the hazards to hearing posed by non-Gaussian noise environments. [DOI: 10.1121/1.1582446]

PACS numbers: 43.66.Ed, 43.50.Pn, 43.50.Qp [NFV]

I. INTRODUCTION

The continuum of noise exposures that produce hearing loss runs from high level continuous Gaussian (G) noise through to military type impulse noise. Studies of noise-induced hearing loss (NIHL) using animal models have most often used either various bandwidths of continuous G noise or impact/impulse transients as the trauma-producing stimulus. Few experimental data are available on the effects of the infinite types of noise that lie between these two extremes, i.e., non-Gaussian (non-G) continuous noise. Industrial noise environments are often non-G, nonstationary and contain high-level noise bursts and/or impact sources.

Hamernik *et al.* (1974) combined high-level noise impulses of the Friedlander type with a G continuous noise to produce a continuous non-G signal. Chinchillas exposed to this noise showed a disproportionate increase in permanent threshold shift (PTS) and sensory cell loss than that produced by either the continuous or impulse noise alone. This potentiation of trauma was eliminated when the impulse was presented in a 2-s window of silence in the otherwise continuous G noise. More recently, Lei *et al.* (1994) showed that continuous non-G noise exposures would produce more trauma than energy and spectrally equivalent G exposures. The non-G character of the noise was produced by the insertion of high-level noise bursts or impact transients into the otherwise G noise. They showed that the increased trauma was related to the kurtosis, $\beta(t)$, of the non-G signal. The systematic relation between trauma metrics and $\beta(t)$ was also shown to be reflected in the spectral distribution of energy in the transients that gave the noise its non-G character (Hamernik and Qiu, 2001). That is, for a fixed long-term spectrum and energy and a constant $\beta(t)$, trauma was directly related to the bandwidth of the transient stimulus. They also showed that over a limited range of $\beta(t)$ the frequency specific kurtosis $\beta(f)$, computed on the filtered noise signal, was well correlated with the additional frequency specific OHC loss produced by the non-G noise relative to the G control condition. These and other animal model experiments (Dunn *et al.*, 1991; Lataye and Campo, 1996) as well as industrial epidemiological data (Thiery and Meyer-Bisch, 1988) suggest limitations on the use of energy-based metrics such as

[a]Electronic mail: roger.hamernik@plattsburgh.edu

TABLE I. Summary of exposure conditions

Group no.	Noise type[a]	N	β(t)	Peak range (dB SPL)	Impulse probability	L_b-background L_{eq} (dBA)
43	(1)	16	3	-	-	-
44	(2)	11	25	[115, 128]	0.6	95.5
51	(2)	12	75	[118, 130]	0.6	91
52	(2)	11	53	[123, 127]	0.6	92.6
53	(2)	12	61	[117, 130]	0.2	94
54	(2)	11	25	[115, 128]	[b]	94
60	(2)	16	39	[115, 129]	0.6	93.5
63	(2)	11	95	[116, 128]	0.6	0
64	(2)	12	12	[116, 126]	0.6	97
49	(3)	12	33	[115, 129]	0.6	91.5
55	(3)	12	25	[115, 129]	0.6	94.5
65	(3)	12	105	[127, 132]	0.1	0
66	(3)	12	15	[113, 127]	0.6	96
68	(3)	11	65	[128, 133]	0.1	96
70	(4)	12	27	[104, 115]	0.6	91.3
71	(3)	12	55	[128, 133]	0.6	0
72	(3)	12	27	[125, 130]	0.1	96.5

[a]Noise type: (1) Broadband Gaussian. (2) Non-Gaussian noise, with or without a Gaussian background component defined by L_b. The transients which provide the non-Gaussian character of the noise were impacts created from three 400-Hz bands of energy centered at 1, 2, and 4 kHz. (3) Same as in (2) except the transients are broadband (710–5680 Hz for group 49 and 1–10 kHz for groups 55, 65, 66, 68, 71, and 72). (4) For this exposure the transient was a broadband (710–5680 Hz) noise burst (100 ms) having a rms SPL between 104 and 115 dB.

[b]Impacts occur regularly 1/1.5 s.

the Leq which are the foundation of current international standards.

The data presented here extend the results of Lei *et al.* (1994) and Hamernik and Qiu (2001) by (1) extending the range of β(t) for each of two series of exposures whose transients have different spectra, and (2) varying the probability of the occurrence of a transient. β(t) was varied by changing both the interval and peak histograms of the noise signal.

II. METHODS

Two hundred and seven (207) chinchillas (between 1 and 2 years old), randomly distributed among 17 exposure groups, were used as subjects. Each animal was made monaural by the surgical destruction, under anesthesia, of the left cochlea. During this procedure a bipolar electrode was implanted, under stereotaxic control, into the left inferior colliculus and the electrode plug cemented to the skull for the recording of auditory evoked potentials (AEP). The AEP was used to estimate pure tone thresholds and surface preparations of the organ of Corti were used to estimate the inner and outer hair cell (IHC, OHC) populations. Additional details of the experimental methods, beyond those presented below, may be found in Ahroon *et al.* (1993).

A. Threshold testing

AEP audiograms were measured at octave intervals from 0.5 to 16.0 kHz. The mean (in dB SPL) of three threshold determinations measured on different days defined each animal's pre- and 30-d post-exposure audiogram. A complete audiogram was measured once daily during each of the 5 days of the exposures and the average (in dB SPL) taken over the 5 days established the mean asymptotic threshold levels and shifts.

B. Histology

Following the last AEP test protocol, each animal was euthanized under anesthesia and the right auditory bulla removed and opened to gain access to the cochlea for perfusion. Fixation solution consisting of 2.5% glutaraldehyde in veronal acetate buffer (final pH=7.3) was perfused through the cochlea. After 12 to 24 h of fixation the cochlea was postfixed in 1% OsO_4 in veronal acetate buffer. Surface preparation mounts of the entire organ of Corti were prepared and IHC, OHC populations were plotted as a function of frequency and location using the frequency-place map of Eldredge *et al.* (1981). Missing cells were identified by their characteristic phalangeal scars. For purposes of this presentation, sensory cell population data is presented as group averages (in percent or numbers missing) taken over octave-band lengths of the cochlea centered on the primary AEP test frequencies.

C. Noise exposures

The noise field was monitored with a Larson Davis 814 sound level meter equipped with a $\frac{1}{2}$-in. microphone. The acoustic signal produced by the Electro-Voice Xi-1152/94 speaker system was transduced by a Brüel and Kjær $\frac{1}{2}$-in. microphone (Model 4134), amplified by a Brüel and Kjær (Model 2610) measuring amplifier and fed to a Windows PC-based analysis system. The design and digital generation of the acoustic signal is detailed in Hsueh and Hamernik (1990, 1991).

During exposure, individual chinchillas were confined to cages ($10\times11\times16$ in.^{3}) with free access to food and water. The 17 groups of animals were exposed to one of the exposure protocols outlined in Table I. The exposure field was uniform to within 2 dB. The exposures lasted 24-h/day for 5

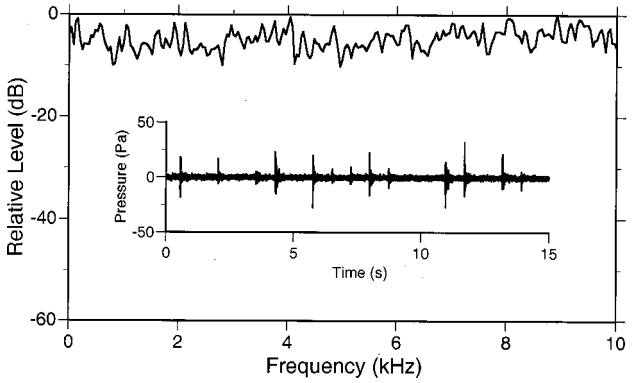


FIG. 1. The average spectrum of a 40-s sample of one of the noise signals. This long-term spectrum was common to all 17 noise exposures. Inset: A 15-s sample of one of the non-Gaussian waveforms. The peak SPL was randomly varied and the probability of an impact occurring in a 750-ms window was 0.6.

days and were interrupted once daily for approximately 20–30 min. for AEP testing. The 5-day continuous exposures produced an asymptotic threshold shift (ATS).

Each exposure had in common approximately the same flat spectrum between 0.125 and 10.0 kHz shown in Fig. 1 and was presented at an Leq = 100 dB(A). This level was chosen so that the G exposure condition would produce consistent but relatively small permanent changes in hearing and sensory cell populations. The 17 exposures differed only in their temporal structures, which were designed to produce 1 G and 16 non-G exposure conditions. The non-G conditions were designed in the frequency domain as described by Hsueh and Hamernik (1990, 1991) and were the result of inserting impacts or noise bursts (group 70), whose spectra were complementary to the background G noise, into the otherwise G signal. Two classes of non-G signals that incorporated impacts were generated. One was produced by inserting impact transients that were generated from three 400-Hz bands of energy centered at 1, 2, and 4 kHz. The other class was produced using broadband impacts that derived their energy from the 710–5680-Hz or 1–10-kHz region of the spectrum. The former are referred to as the ''three band'' and the latter as the ''broadband'' non-G exposures. The transient peak levels were randomly varied between the limits indicated in Table I and the probability (P) of an impact or noise burst occurring in a 750-ms window was set at either 0.6 or 0.1 except for groups 53 and 54. Peak and interval histograms, which varied for each condition, were constructed for each of the exposure conditions and β(t) and β(f) were computed. β(f) was computed on octave bandwidths of the filtered signal centered at the audiometric test frequencies to provide, in effect, a ''kurtosis spectrum.''

III. RESULTS AND DISCUSSION

A. Noise analysis

Figure 1 shows the average spectrum (relatively flat, broadband) obtained from a 40-s sample of the digitized waveform. This long-term spectrum was common to each of the 17 different exposure conditions. The inset shows a 15-s sample of one of the non-G, nonstationary waveforms. Figures 2(a) and (c) show the spectra of the ''three band'' and ''broadband'' impact transients that produced the non-G signals along with a sample waveform. A second type of broadband (0 to 10 kHz) transient stimulus, whose spectrum is not shown, was used for the exposure of groups 55, 65, 66, 68, 71, and 72. Figures 2(b) and (d) show typical complementary spectra of the G components of the two broad classes of non-G signals. Table II presents the octave band and overall levels of each noise exposure. Values shown are the mean values obtained from eight 40-s samples of the digitized waveform.

Amplitude histograms for four of the non-G exposures are shown in Figs. 3(a)–(d). These were constructed

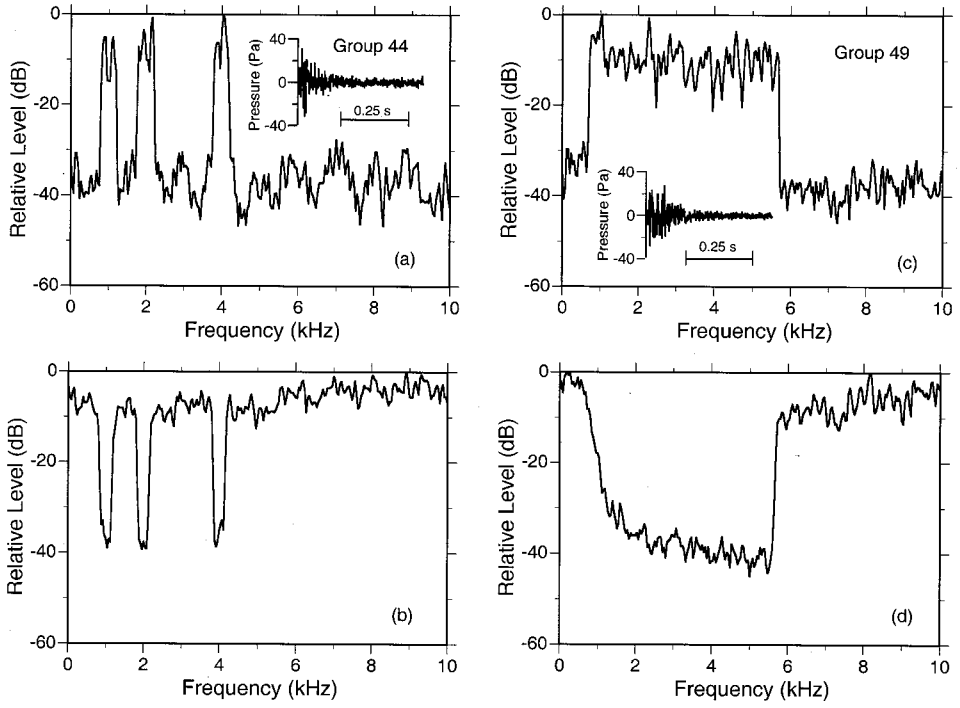


FIG. 2. The upper panels show the spectrum of (a) one of the ''three band'' impacts that was created from three 400-Hz bands of energy centered on 1.0, 2.0, and 4.0 kHz and (c) one of the ''broadband'' impacts that derived its energy from the 710–5680-Hz region of the spectrum. The insets show sample impact waveforms. Panels (b) and (d) show the respective complementary spectra of the Gaussian noise which was mixed with the impacts to create the non-Gaussian noise exposures. Group numbers 44 and 49 identify the specific exposures outlined in Table I.

TABLE II. Total and octave band sound pressure levels (dB SPL) over the 5-day exposure period for all experimental groups.

Group no.	Octave band cf (kHz) 0.5	1	2	4	8	16	Mean Leq	Mean Leq(A)	s.d.
43	89	89	88	91	99	98	103	100	0.04
44	82	93	95	95	95	95	101	100	0.26
49	86	95	94	96	93	93	101.5	101	0.65
51	84	97	93	93	91	89	101	100	0.89
52	88	98	94	94	93	94	102	101	0.18
53	87	97	93	93	94	94	101	100	0.81
54	86	97	93	93	94	96	102	100	0.40
55	93	93	89	91	97	93	102	100	0.77
60	86	96	92	93	92	93	101	100	0.67
63	85	98	92	92	83	77	100	100	0.44
64	88	96	93	95	96	96	103	101	0.39
65	96	95	89	90	94	84	103	100	0.35
66	91	91	90	94	99	95	103	100	0.17
68	95	94	89	92	97	94	103	101	0.52
70	86	93	93	96	93	94	101	100	0.34
71	95	93	90	94	97	86	103	100	0.45
72	92	91	89	94	98	95	103	100	0.52

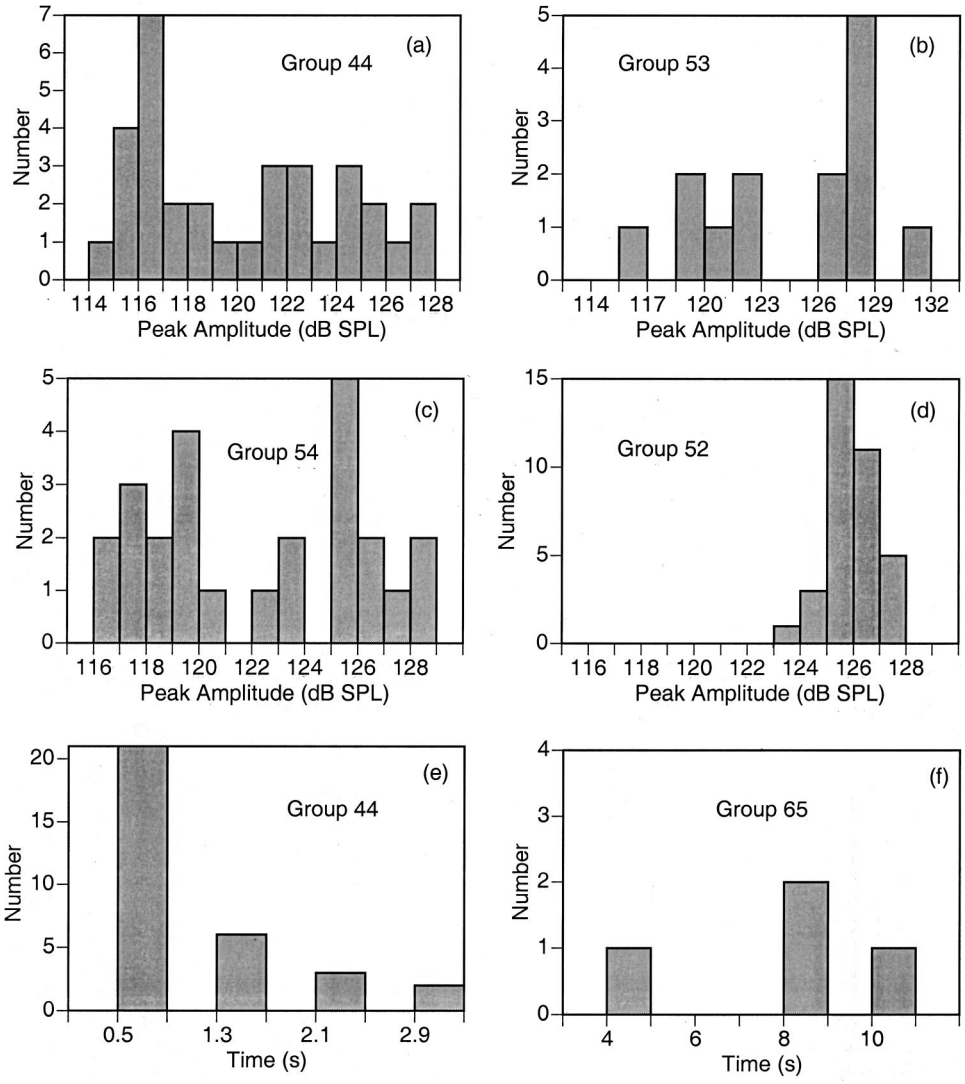


FIG. 3. Panels (a)–(d) show impact peak histograms from a 40-s sample of the exposures identified in each panel. Panels (e) and (f) show impact interval histograms for impact probabilities of 0.6 and 0.1, respectively.

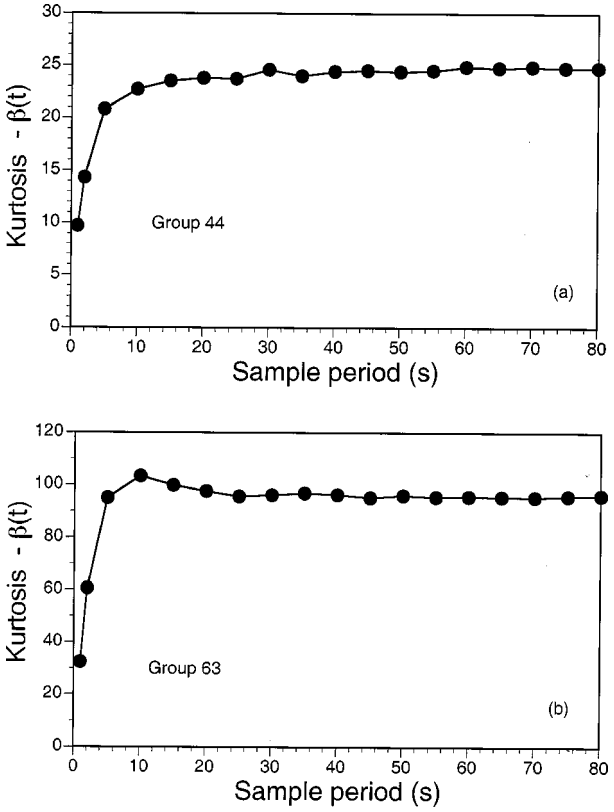


FIG. 4. The kurtosis β(t) computed on sample windows varying from 1 to 80 s for two of the non-Gaussian exposure groups 44 and 63. β(t) reaches a relatively stable value for windows greater than about 30 s.

from a 40-s sample of the noise. Since the peak levels were randomly generated, each sample of the signal will yield a different histogram. These histograms, however, illustrate that the net effect of randomizing the peaks is to produce very different distributions for each of the non-Gaussian exposures. Note that for group 52 the peak SPLs are, by design, nearly constant. The peak interval histograms for two exposure conditions illustrated in Figs. 3(e) and (f) reflect the 0.6 and 0.1 probability of an impact occurring in a 750-ms window respectively. Note that for group 54 the impacts were presented at a constant rate of 1/1.5 s and that the probability

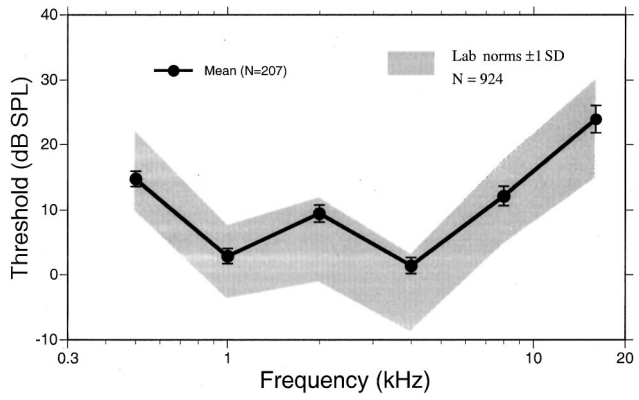


FIG. 5. The mean ($N=207$) AEP audiogram for the animals used in the various exposure conditions outlined in Table I compared to the laboratory average based on $N=924$ animals. The shaded region represents ± one standard deviation of the laboratory standard. Bars on the data points represent one standard error of the mean.

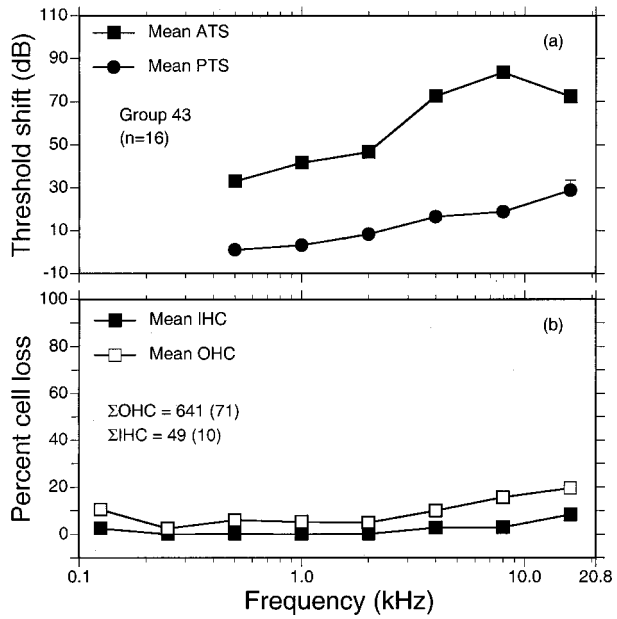


FIG. 6. (a) The group mean asymptotic and permanent threshold shifts (ATS, PTS) for group 43 exposed to the Gaussian noise. ATS is the average of the daily AEP threshold estimates over the 5-day exposure. PTS is the average of three threshold estimates taken 30 days postexposure. (b) The group mean percent outer and inner hair cell (OHC, IHC) loss averaged over adjacent octave band lengths of the basilar membrane. Bar=standard error. Total average numbers of IHCs and OHCs missing and standard errors are also given.

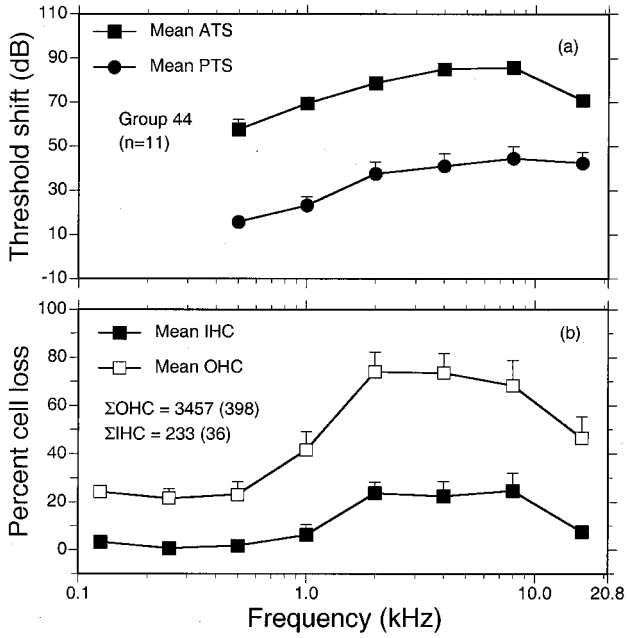


FIG. 7. (a) The group mean asymptotic and permanent threshold shifts (ATS, PTS) for group 44 exposed to a non-Gaussian noise. ATS is the average of the daily AEP threshold estimates over the 5-day exposure. PTS is the average of three threshold estimates taken 30 days postexposure. (b) The group mean percent outer and inner hair cell (OHC, IHC) loss averaged over adjacent octave band lengths of the basilar membrane. Bars =standard errors. Total average numbers of IHCs and OHCs missing and standard errors are also given.

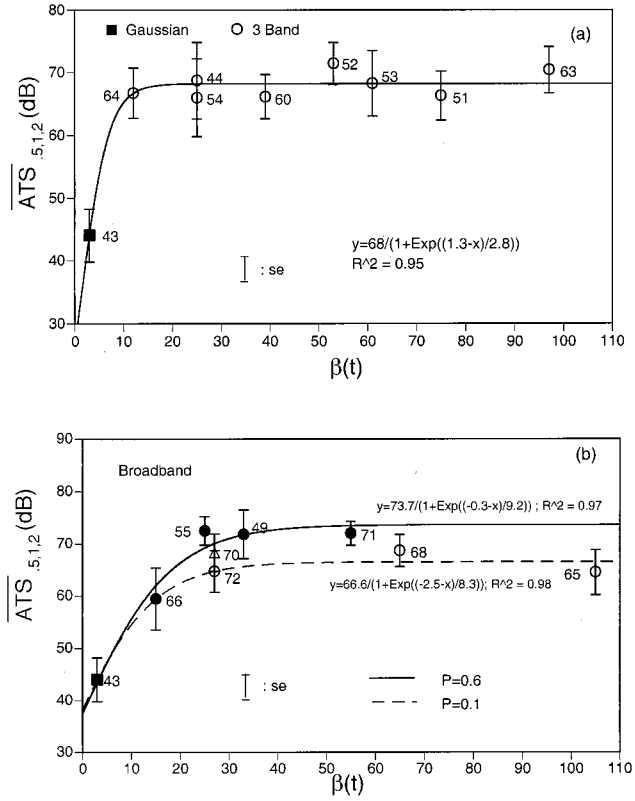


FIG. 8. The group mean asymptotic threshold shift averaged over the 0.5, 1.0, and 2.0 kHz AEP test frequencies ($\overline{ATS}_{.5,1,2}$) as β(t) is increased for (a) the ''three band'' class of non-Gaussian noise exposures (○) and (b) the ''broadband'' class of non-Gaussian noise exposures (○, $P=0.1$; ●, $P=0.6$). ■=the Gaussian reference exposure. Bars=standard errors. The nonlinear regression equation and the coefficient of determination (r^2) are given.

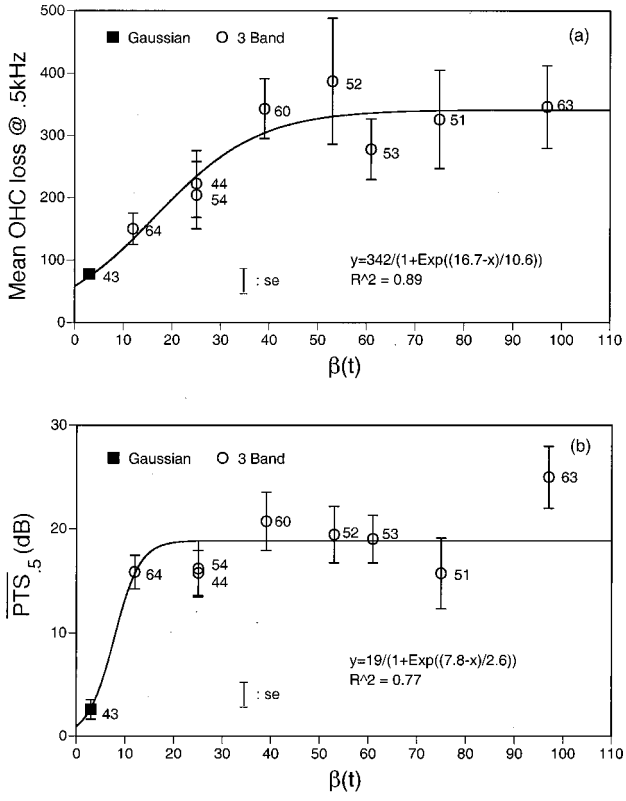


FIG. 9. (a) The group mean number of outer hair cells (OHC) lost in the 0.5-kHz octave band region of the basilar membrane and (b) the group mean permanent threshold shift measured at the 0.5-kHz test frequency ($\overline{PTS}_{.5}$) as β(t) is increased. ○=the ''three band'' class of non-Gaussian noise exposures. ■=the Gaussian reference condition. Bar=standard error. The nonlinear regression equation and the coefficient of determination (r^2) are given.

of an impact occurring in group 53 was set at 0.2.

The kurtosis β of the sample distribution is defined as the ratio of the fourth-order central moment to the squared second-order central moment of the distribution. β is sensitive to the various parameters that define a complex noise such as the levels and durations of the transients, temporal structure of the noise, the crest factor, etc., as well as to the duration of the noise sample over which β is computed. The values of β presented in Table I are an average computed over eight 40-s sample of the noise. Figure 4 shows the typical asymptotic behavior of β for two of the exposure conditions as the sample duration is increased. For all the noise conditions used, β reached an approximately constant value for a sample window duration of approximately 40 s or more.

B. Audiometric and histological effects of exposure

Mean AEP preexposure thresholds for all animals ($N=207$) are compared to the laboratory standard ($N=924$) in Fig. 5. Standard errors were small and all animals and groups fell within ± one standard deviation of laboratory norms and reported thresholds for the chinchilla (Fay, 1988). ANOVA analysis indicated no statistically significant effect of group. In this and all other figures bars denote the standard error. Where a bar is not shown, the standard error was less than the size of the datum symbol.

Figures 6 and 7 show examples of complete data sets for two of the exposure conditions; group 43, the G, $\beta(t)=3$ reference condition, and group 44, a non-G, $\beta(t)=25$ condition. The upper panels present the group mean ATS and PTS while the lower panels show the corresponding cochleograms. Total OHC and IHC losses are also presented along with the standard errors. These two figures illustrate one of the consistent findings to emerge from all of these exposures. Specifically, for an equivalent spectral distribution of energy and total energy, the non-G exposures produce more ATS, PTS and sensory cell loss than does the energy equivalent G exposure. Similar data summaries were prepared for the other 15 exposure conditions. It is from data files such as these that all the remaining figures are derived.

Figure 8 shows the mean ATS averaged over the 0.5-, 1.0-, and 2.0-kHz test frequencies ($\overline{ATS}_{.5,1,2}$) for the ''three band'' (upper panel) and the ''broadband'' (lower panel) non-G exposures. (Note: The ''three band'' and 'broadband' designations in this and remaining figures refer to the spectral content of the transients that produced the non-G noise exposures. The numbers next to the datum points in this and subsequent figures refer to exposure conditions shown in Table I.) The lower test frequencies were chosen for this average because the ATS levels at the higher test frequencies, for the non-G conditions, reached or exceeded the upper limits of our AEP testing system. For all exposure conditions

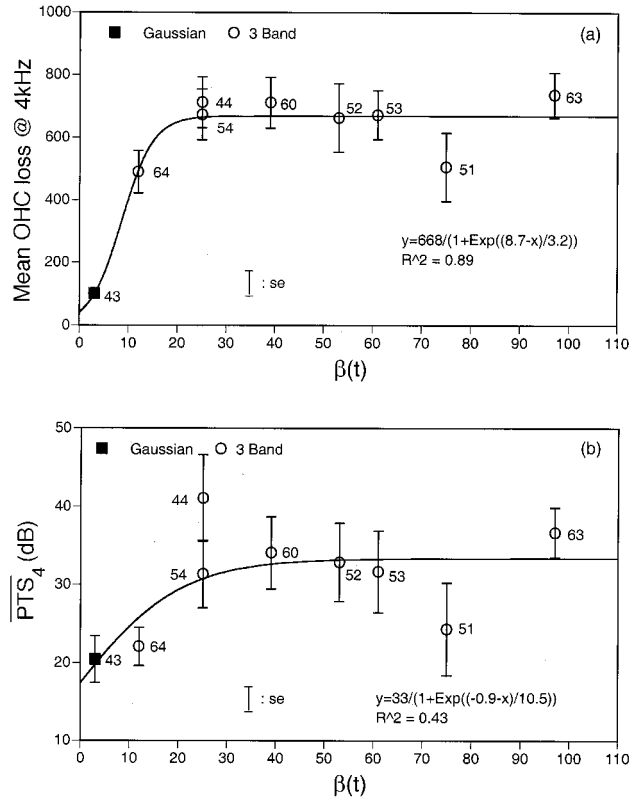


FIG. 10. (a) The group mean number of outer hair cells (OHC) lost in the 4.0-kHz octave band region of the basilar membrane and (b) the group mean permanent threshold shift measured at the 4.0-kHz test frequency ($\overline{\text{PTS}}_4$) as β(t) is increased. ○ = the "three band" class of non-Gaussian noise exposures. ■ = the Gaussian reference condition. Bar = standard error. The nonlinear regression equation and the coefficient of determination (r^2) are given.

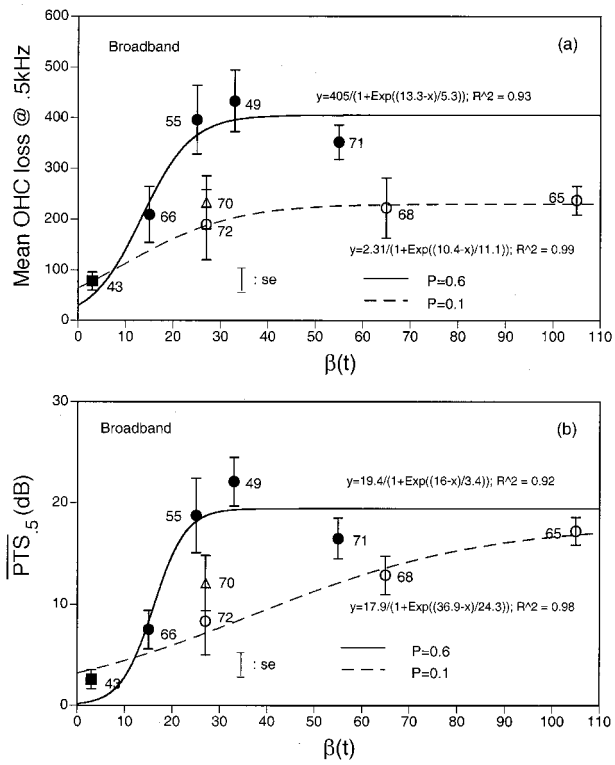


FIG. 11. (a) The group mean number of outer hair cells (OHC) lost in the 0.5-kHz octave band region of the basilar membrane and (b) the group mean permanent threshold shift measured at the 0.5-kHz test frequency ($\overline{\text{PTS}}_{.5}$) as β(t) is increased for the "broadband" class of non-Gaussian noise exposures. ●, $P=0.6$; ○, $P=0.1$; ■ = the Gaussian reference condition. Bar = standard error. The nonlinear regression equations and the coefficients of determination (r^2) are given.

shown in Fig. 8(a), $\overline{\text{ATS}}_{.5,1,2}$ exceeded that produced by the G noise (group 43) by 20 dB or more and was relatively constant across the eight different non-Gaussian exposure conditions regardless of the statistical properties of the non-G noise. The exposure of group 63 (the end point of this sequence of exposures) consisted of a sequence of random transients of the type shown in Figs. 2(a) and (b), i.e., there was no background G component ($L_b=0$). Thus, although group 63 has a different spectrum, it represents the end of a continuum of exposures that start with the G exposure of group 43. Note that in order to increase β(t) the level of the G component of the non-G exposures was gradually reduced from a high of 97 dB for group 64, to 91 dB for group 51, and finally to zero for group 63, i.e., the crest factor was increased. A similar presentation of data is shown in Fig. 8(b) for the "broadband" conditions and for two different probabilities ($P=0.6$ and $P=0.1$) for generating the transients in the non-G noise. For the "broadband" non-G exposures, $\overline{\text{ATS}}_{.5,1,2}$ gradually increased as β(t) increased and appeared to reach an upper limit at a β(t) of approximately 40, of about 75 dB for the $P=0.6$ condition and 65 dB for the $P=0.1$ condition. This is about 30 and 20 dB respectively greater than that produced by the G condition. Note also in this series of exposures that groups 71 and 65 were exposed to pure impact noise.

The permanent effects of these exposures, quantified by PTS and OHC loss, are presented in Figs. 9–14. Figures 9–12 [(a) panels] show OHC losses in the octave band length of the basilar membrane centered at 0.5 and 4.0 kHz, while the (b) panels show the corresponding PTS for the "three band" (Figs. 9 and 10) and "broadband" (Figs. 11 and 12) conditions. As seen with the ATS data, there is a consistent increase in the PTS of up to 35 dB and very large increases in OHC loss for the non-G conditions. Generally the greatest differences between the G and non-G conditions were found in the "broadband" groups for $\beta(t)>30$. This result is in agreement with the data of Hamernik and Qiu (2001). The asymptotic behavior of both PTS and OHC loss for $\beta(t)>\sim 40$ is also reflected in the frequency specific permanent losses. In the "broadband" conditions the exposures with $P=0.1$ consistently showed less permanent effects while in the "three band" conditions groups 53 with $P=0.2$ and 54 with impacts occurring at a constant rate of 1/1.5 s produced permanent effects that were similar to the $P=0.6$ exposure conditions.

Figures 13 and 14 show the group mean total number of OHCs lost and the mean PTS computed at 2.0, 4.0, and 8.0 kHz ($\overline{\text{PTS}}_{2,4,8}$) for all exposure conditions. These frequencies were chosen for an average index of trauma since they generally showed the most PTS. There is a consistent relation between β(t) and the number of missing OHCs. In the "three band" non-G exposures, OHC loss exceeded that found fol-

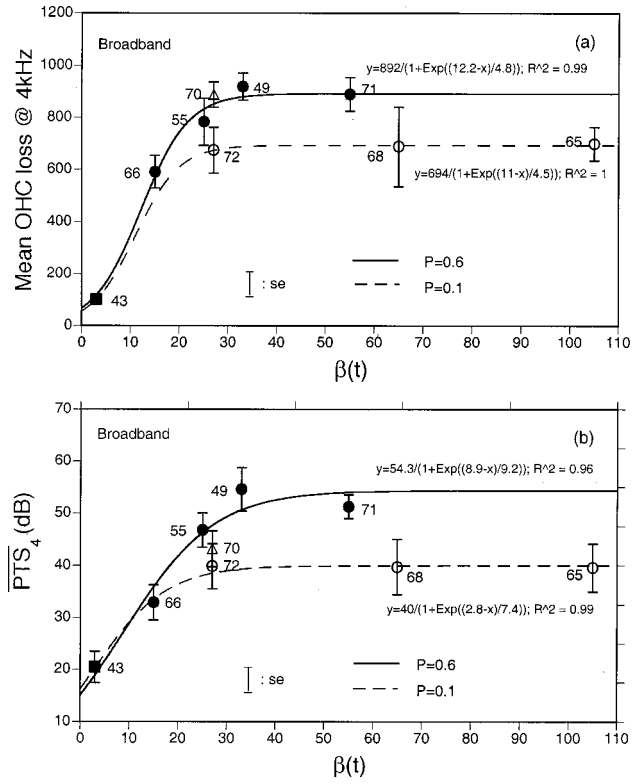


FIG. 12. (a) The group mean number of outer hair cells (OHC) lost in the 4.0-kHz octave band region of the basilar membrane and (b) the group mean permanent threshold shift measured at the 4.0-kHz test frequency ($\overline{\text{PTS}_4}$) as β(t) is increased for the ''broadband'' class of non-Gaussian noise exposures. ●, $P=0.6$; ○, $P=0.1$; ■=the Gaussian reference condition. Bar =standard error. The nonlinear regression equations and the coefficients of determination (r^2) are given.

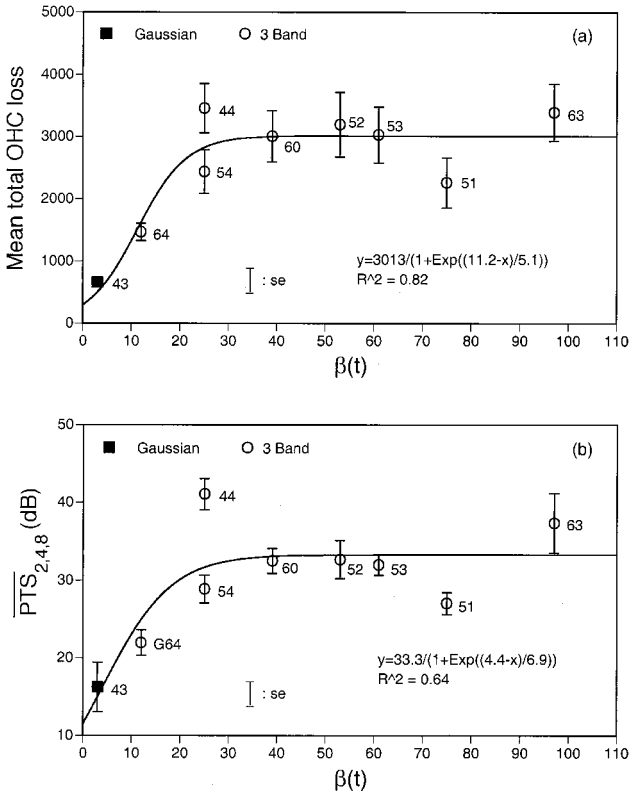


FIG. 13. (a) The group mean total number of outer hair cells (OHC) lost and (b) the group mean permanent threshold shift averaged at the 2.0-, 4.0-, and 8.0-kHz test frequencies ($\overline{\text{PTS}_{2,4,8}}$) as β(t) is increased. ○=the ''three band'' class of non-Gaussian noise exposures. ■=the Gaussian reference condition. Bar=standard error. The nonlinear regression equation and the coefficient of determination (r^2) are given.

lowing the G exposure by a factor of 5. The increased OHC loss was even greater for the ''broadband'' conditions. $\overline{\text{PTS}_{2,4,8}}$ in the non-G conditions exceeded that produced by the G condition by 20 to 35 dB depending on the spectral features of the noise. There were no differences in the permanent effects among exposure groups 60, 52, and 53. For group 53 the peaks were random and the probability of occurrence was 0.2, while for exposure of group 52 with $P=0.6$, the peaks were kept relatively constant [Fig. 3(d)]. Note also that in Fig. 14 the level of trauma sustained by group 70 from a non-G exposure that incorporated noise bursts having $P=0.6$ instead of impacts was considerably reduced compared to group 55 that had similar statistical properties but incorporated impacts. The trauma in group 70 fell roughly near the ''broadband'' impact conditions with $P=0.1$. The asymptotic behavior of the regression lines in Figs. 13 and 14 would indicate that, for the chinchilla model, under our limited exposure conditions, the use of an energy principal as proposed a number of years ago by Martin (1976) and others may be useful for the evaluation of non-G noise but only for $\beta(t)>40$.

Figure 15 illustrates, for three of the exposure conditions, the relation between the β(f) spectrum and the frequency specific excess OHC loss produced by the non-Gaussian exposures (i.e., the OHC loss produced in each octave band by the non-Gaussian minus that produced by the Gaussian noise). In agreement with the results of Lei *et al.* (1994) and Hamernik and Qiu (2001), these examples illustrate the relatively close relation between β(f) and the additional OHC loss produced by non-Gaussian noise exposures.

IV. SUMMARY AND CONCLUSIONS

Under a variety of exposure conditions the results presented here agree with our earlier results (Hamernik and Qiu, 2001; Lei *et al.*, 1994) in showing the following: (1) Non-G noise exposures produce more hearing trauma than do energy and spectrally equivalent G noise exposures. This is true whether the transients that produce the non-G character of the noise are impacts or noise bursts (e.g., group 70). (2) The statistical metric, kurtosis (β), in conjunction with an energy metric can identify hazardous exposure conditions not identified by conventional energy based metrics alone. In the chinchilla model, PTS and OHC loss are monotonically related to β(t) over the range $3<\beta(t)<40$. (3) The extent of trauma is related to the bandwidth of the transients that give the noise its non-G character. (4) A frequency specific kurtosis β(f) can be computed on the filtered acoustic signal. β(f) is well correlated with the frequency-specific OHC loss produced by non-G noise exposures relative to the energy and spectrally equivalent G exposures.

The results in this paper also extend our earlier results by showing the following:

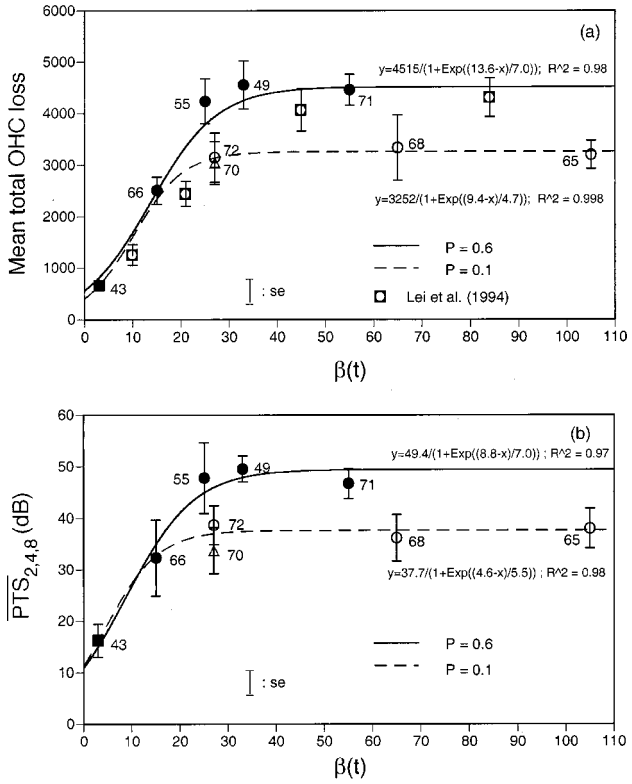


FIG. 14. (a) The group mean total number of outer hair cells (OHC) lost and (b) the group mean permanent threshold shift averaged at the 2.0-, 4.0-, and 8.0-kHz test frequencies ($\overline{\text{PTS}}_{2,4,8}$) as β(t) is increased for the ''broadband'' class of non-Gaussian noise exposures. ●, $P=0.6$; ○, $P=0.1$; ■=the Gaussian reference condition. Bar=standard error. The nonlinear regression equations and the coefficients of determination (r^2) are given.

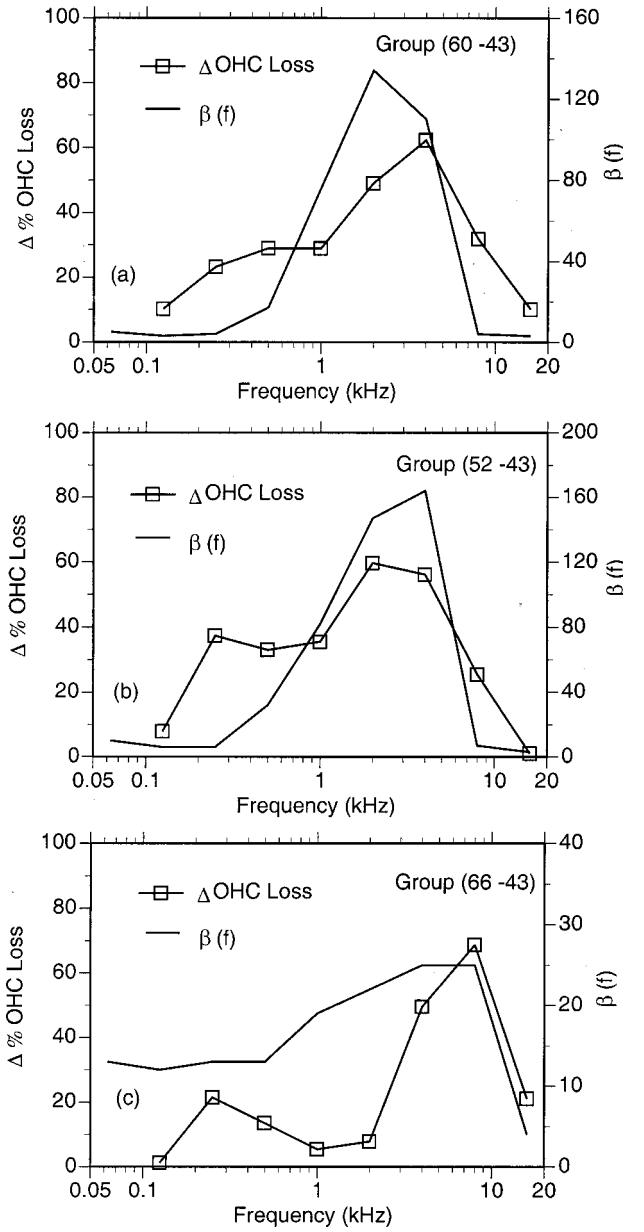


FIG. 15. Three examples of the relation between the increased octave band percent outer hair cell loss produced by the non-Gaussian exposure relative to the energy equivalent Gaussian exposure and the ''kurtosis spectrum'' β(f). Δ%OHC loss=the percent OHC loss produced in a given octave band region by the non-Gaussian noise minus that produced by the Gaussian noise.

(1) For $\beta(t)>\sim 40$ and a fixed probability (P) of a transient occurring, the degree of trauma as measured by ATS, PTS, or OHC loss remains constant despite changes in the statistical characteristics of the noise exposure. That is, varying the amplitude or the interval histogram of the transients had no effect on the indices of trauma. These variables are incorporated in the kurtosis metric. The function relating trauma to β(t) reaches an asymptote for $\beta(t)>\sim 40$. The asymptotic behavior of the function relating the indices of trauma to β(t) for $P=0.6$ and $P=0.1$ suggests that, over a range of values of the kurtosis, an energy-based approach can be used to assess the potential of an exposure to produce trauma. This result supports the approach in some standards documents (e.g., ISO-1999, 1990) that assign a ''correction'' factor for noise environments that contain impacts. There does not appear, however, to be experimental justification for what the value of the factor should be.

(2) Decreasing the probability of a transient occurring from $P=0.6$ to $P=0.1$ reduced the degree of trauma for the exposures that incorporated broadband impacts. This would be expected since in the limit as P goes to zero, the exposure approaches the G condition. However, for the ''3 band'' series of exposures neither the $P=0.2$ (group 53) nor the constant impact presentation rate (group 54) showed a level of trauma that differentiated these groups from those with exposures having $P=0.6$.

(3) One of the exposures (group 70) used random noise bursts instead of impacts to produce the non-G exposure. The trauma produced by this exposure, while greater than that produced by the G noise, was significantly less than that which was produced by the exposure conditions for group 55. The exposures for groups 55 and 70 had approximately the same β(t) and differed only in the envelope of the transient and L_b. Although the envelope of the transient and its rise time are incorporated in β(t), β(t) alone could not be used to differentiate between these exposures. However, when data from two groups that used noise burst transients to produce non-G noise in the Lei *et al.* (1994) paper are compared with the results of group 70 [Fig. 14(a)], a trauma versus β(t) func-

tion emerges that parallels the function derived from the impact data. Specifically, in the Lei *et al.* paper, groups CN VII with $\beta(t)=21$ and CN IX with $\beta(t)=10$ showed an increased trauma relative to the G condition as seen with our group 70 with $\beta(t)=27$. The data from these three conditions are well ordered with respect to $\beta(t)$. This suggests that ''correction'' factors in standards need to be used not only for impact containing noise environments but also for any non-G environment whether or not the transients are impacts. Note further that for the three exposures identified above there is a linear relation between trauma and $\beta(t)$ with, as in the case of impacts, trauma increasing with increasing $\beta(t)$. For these exposures having $\beta(t)<40$, any asymptotic behavior of the trauma versus $\beta(t)$ function could not be determined. Note also that the data from the two exposure conditions in the Lei *et al.* paper that employed impacts (i.e., Groups CN VI and IX) and whose statistical properties were different from those used in the present paper, fall along the asymptotic line determined by the $P=0.6$ exposure conditions (Fig. 14).

In conclusion, the data from the abundant and varied exposure conditions show that the kurtosis of the amplitude distribution of a noise environment is an important variable in determining the hazards to hearing posed by non-Gaussian noise environments.

ACKNOWLEDGMENTS

This work was supported by Grant No. 1-R01-OH02317 from the National Institute for Occupational Safety and Health. The able technical assistance of George A. Turrentine, Ann Johnson, and Diane Fresch is greatly appreciated. In conducting the research described in this study, the investigators adhered to the ''Guide for the Care and Use of Laboratory Animals,'' prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council [DHHS Publication No. (NIH) 86-23, revised 1985].

Ahroon, W. A., Hamernik, R. P., and Davis, R. I. (**1993**). ''Complex noise exposures: An energy analysis,'' J. Acoust. Soc. Am. **93**, 997–1006.

Dunn, D. E., Davis, R. R., Merry, C. J., and Franks, J. R. (**1991**). ''Hearing loss in the chinchilla from impact and continuous noise exposure,'' J. Acoust. Soc. Am. **90**, 1979–1985.

Eldredge, D. H., Miller, J. D., and Bohne, B. A. (**1981**). ''A frequency-position map for the chinchilla cochlea,'' J. Acoust. Soc. Am. **69**, 1091–1095.

Fay, R. A. (**1988**). *Hearing in Vertebrates* (Hill-Fay Associates, Winnetka, IL).

Hamernik, R. P., and Qiu, W. (**2001**). ''Energy-independent factors influencing noise-induced hearing loss in the chinchilla model,'' J. Acoust. Soc. Am. **110**, 3163–3168.

Hamernik, R. P., Henderson, D., Crossley, J. J., and Salvi, R. J. (**1974**). ''Interaction of impulse and continuous noise: audiometric and histological effects,'' J. Acoust. Soc. Am. **55**, 117–121.

Hsueh, K. D., and Hamernik, R. P. (**1990**). ''A generalized approach to random noise synthesis: Theory and computer simulation,'' J. Acoust. Soc. Am. **87**, 1207–1217.

Hsueh, K. D., and Hamernik, R. P. (**1991**). ''Performance characteristics of a phase domain approach to random noise synthesis,'' Noise Control Eng. J. **36**, 18–32.

ISO-1999 (**1990**). International Organization for Standardization, ''Acoustics: Determination of occupational noise exposure and estimation of noise-induced hearing impairment,'' ISO-1999, Geneva.

Lataye, R., and Campo, P. (**1996**). ''Applicability of the L_{eq} as a damage-risk criterion: An animal experiment,'' J. Acoust. Soc. Am. **99**, 1621–1632.

Lei, S-F., Ahroon, W. A., and Hamernik, R. P. (**1994**). ''The application of frequency and time domain kurtosis to the assessment of hazardous noise exposures,'' J. Acoust. Soc. Am. **96**, 1435–1444.

Martin, A. (**1976**). ''The equal energy concept applied to impulse noise,'' in *Effects of Noise on Hearing*, edited by D. Henderson, R. P. Hamernik, D. S. Dosanjh, and J. Mills (Raven Press, New York), pp. 421–453.

Thiery, L., and Meyer-Bisch, C. (**1988**). ''Hearing loss due to partly impulsive industrial noise exposure at levels between 87 and 90 dB (A),'' J. Acoust. Soc. Am. **84**, 651–659.

Intensity discrimination and increment detection in cochlear-implant users[a)]

Magdalena Wojtczak[b)]
Psychoacoustics Laboratory, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

Gail S. Donaldson
Clinical Psychoacoustics Laboratory, University of Minnesota, 420 Delaware Street SE, Minneapolis, Minnesota 55455

Neal F. Viemeister
Psychoacoustics Laboratory, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 19 June 2002; revised 4 April 2003; accepted 14 April 2003)

Intensity difference limens (DLs) were measured in users of the Nucleus 22 and Clarion v1.2 cochlear implants and in normal-hearing listeners to better understand mechanisms of intensity discrimination in electric and acoustic hearing and to evaluate the possible role of neural adaptation. Intensity DLs were measured for three modes of presentation: gated (intensity increments gated synchronously with the pedestal), fringe (intensity increments delayed 250 or 650 ms relative to the onset of the pedestal), and continuous (intensity increments occur in the presence of a pedestal that is played throughout the experimental run). Stimuli for cochlear-implant listeners were trains of biphasic pulses; stimuli for normal-hearing listeners were a 1-kHz tone and a wideband noise. Clarion cochlear-implant listeners showed level-dependent effects of presentation mode. At low pedestal levels, gated thresholds were generally similar to thresholds obtained in the fringe and continuous conditions. At higher pedestal levels, however, the fringe and continuous conditions produced smaller intensity DLs than the gated condition, similar to the gated-continuous difference in intensity DLs observed in acoustic hearing. Nucleus cochlear-implant listeners did not show consistent threshold differences for the gated and fringe conditions, and were not tested in the continuous condition. It is not clear why a difference between gated and fringe thresholds occurred for the Clarion but not the Nucleus subjects. Normal-hearing listeners showed improved thresholds for the continuous condition relative to the gated condition, but the effect was larger for the 1-kHz tonal carrier than for the noise carrier. Findings suggest that adaptation occurring central to the inner hair cell synapse mediates the gated-continuous difference observed in Clarion cochlear-implant listeners and may also contribute to the gated-continuous difference in acoustic hearing. [DOI: 10.1121/1.1579007]

PACS numbers: 43.66.Fe, 43.66.Ts [MRL]

I. INTRODUCTION

Information encoded in the fine structure of speech and other environmental sounds may be unavailable to cochlear-implant users due to the use of pulsatile stimuli by the speech processor. Thus, these listeners with electric hearing may rely strongly on information encoded in the stimulus envelope. To maximize performance of cochlear-implant listeners, it is important to maximize the perceptual resolution in both the intensity/amplitude and temporal domains. A number of studies have measured intensity discrimination in electric hearing. Some have measured intensity discrimination in animals (e.g., Pfingst *et al.*, 1983; Pfingst and Rai, 1990), whereas others have tested human cochlear-implant subjects (Hochmair-Desoyer *et al.*, 1981; Shannon, 1983; Dillier *et al.*, 1983; Busby *et al.*, 1992; Nelson *et al.*, 1996; Donaldson and Viemeister, 2000). All of these existing studies have employed a gated mode of presentation in which intensity increments are equal in duration to the pedestals (see Fig. 1), even though in everyday life, intensity changes are most commonly encountered in ongoing sounds.

Psychophysical studies in listeners with normal hearing have demonstrated that sensitivity to intensity increments depends upon mode of stimulus presentation. Viemeister and Bacon (1988) reported that over a wide range of levels, the Weber fraction ($10 \log \Delta I/I$, where ΔI is the intensity DL), measured with a continuous tonal pedestal, is on average about 4.6 dB smaller than the Weber fraction measured when the pedestal is gated with an increment of the same duration. Their result supported earlier findings that sensitivity to intensity increments is generally better for increments occurring in a continuous sound (Campbell and Lasky, 1967; Green, 1969; Zwicker and Fastl, 1972; Green *et al.*, 1979; Bacon and Viemeister, 1985). More recently, Moore and Peters (1997) studied the effects of gating the pedestal before the onset of the increment. They found an effect consistent in the direction but smaller than the gated-continuous difference reported by Viemeister and Bacon (1988), but the gated (no-fringe) condition was not used in their study.

[a)]Portions of these data were presented at the 141st Meeting of the Acoustical Society of America [M. Wojtczak, G. S. Donaldson, and N. F. Viemeister, J. Acoust. Soc. Am. **109**, 2464(A) (2001)].

[b)]Author to whom correspondence should be addressed. Electronic mail: wojtc@umn.edu

 0001-4966/2003/114(1)/396/12/$19.00

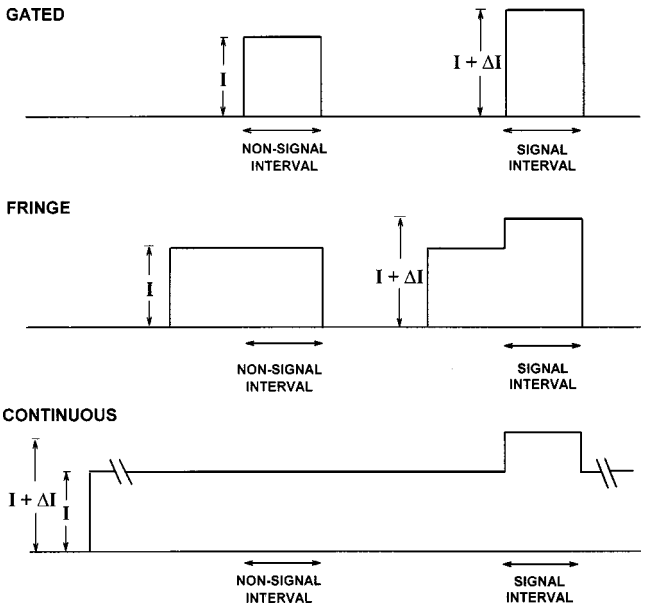


FIG. 1. A schematic illustration of the conditions in which intensity DLs were measured for cochlear-implant users and listeners with normal hearing. The top row shows the gated condition, in which the increment (ΔI) is gated on simultaneously with the pedestal (I); the middle row shows the fringe condition, in which the increment is delayed relative to the onset of the pedestal; and the bottom row shows the continuous condition, in which the pedestal is played continuously throughout the experimental run and the increment occurs during the signal interval.

The most compelling explanations for the gated-continuous difference involve neural adaptation (Smith and Zwislocki, 1975; Smith, 1977). They suggest that adaptation affects increment detection because it effectively decreases the internal response to the pedestal, while having no effect on the incremental response when the signal of a fixed intensity is added at different time delays relative to the onset of the pedestal. Assuming that the increment is detected when the internal response increases by some criterion factor relative to the response just before the increment, adaptation correctly predicts that delaying the onset of the increment will lead to a lower threshold. This explanation is closely related to the idea of a change-detection mechanism proposed by Macmillan (1971, 1973). This mechanism is assumed to be sensitive to signal onsets and offsets but not to be sensitive to the direction of a change. A constant criterion change in response necessary for detection would be in agreement with the above adaptation explanation. It is not clear whether the main source of adaptation is in the synapse between the inner hair cells and the spiral ganglion, as suggested by many physiological studies (Geisler *et al.*, 1979; Schwid and Geisler, 1982; Smith and Brachman, 1982; Meddis, 1986; Westerman and Smith, 1988; Javel, 1996), or if the dominant adaptation originates at a more central site of processing (Shannon and Otto, 1990).

One goal of the present study was to measure cochlear-implant listeners' sensitivity to intensity changes in an ongoing sound. To our knowledge, detection of increments in ongoing stimuli has never been measured in electric hearing. Another goal was to determine if the mechanism underlying the gated-continuous difference observed for normal-hearing listeners has its origin in peripheral or more central processing. Cochlear processing is by-passed in electric hearing. Thus, if neural adaptation underlies the gated-continuous difference in acoustic hearing and it stems primarily from depletion of neurotransmitter at inner hair cell synapses, then there should be no difference between gated and continuous thresholds in cochlear-implant users. If, on the other hand, adaptation underlying the gated-continuous difference originates at a more central site, then cochlear-implant listeners would be expected to exhibit a difference between gated and continuous thresholds. It should be noted that due to the absence of the basilar-membrane nonlinearity, the size of the observed difference might not match that for the normal-hearing listeners.

Two experiments were performed to compare increment detection in an ongoing sound with intensity discrimination for gated stimuli. In the first experiment, sensitivity to intensity increments was measured in cochlear-implant users under two conditions. In one case, the pedestal was gated synchronously with the increment (gated condition), whereas in the other its onset occurred a few hundred milliseconds before the increment (fringe condition). This experiment allowed us to evaluate whether the thresholds reported in various studies of intensity discrimination in electric hearing are representative of the listeners' ability to detect small intensity changes in ongoing stimuli, a situation most commonly encountered in daily life. Normal-hearing listeners were also tested in these two conditions for comparison. In the second experiment, cochlear-implant listeners were tested in the gated condition and in a condition in which the pedestal was played continuously throughout the experimental run (continuous condition). A diagram presenting the three conditions in which intensity DLs were measured is shown in Fig. 1. Our general hypothesis was that cochlear-implant listeners would demonstrate similar intensity DLs in all conditions, based on the assumption that peripheral adaptation underlies the gated-continuous difference in acoustic hearing.

II. EXPERIMENT 1

The purpose of this experiment was to determine whether delaying the onset of an intensity increment relative to the onset of the pedestal would facilitate increment detection. Intensity-discrimination and increment-detection thresholds were measured in cochlear-implant subjects and subjects with normal hearing. Normal-hearing listeners were included because the existing literature does not provide a clear picture of the dependence of intensity DLs on the duration of the preceding fringe.

A. Gated versus fringe conditions in electric hearing

1. Subjects

Four users of the Nucleus-22 cochlear implant (N28, N30, N31, N32) and three users of the Clarion-1.2 cochlear implant (C14, C16, C12) participated in the study. Table I provides information about the listeners, including their age, the duration of hearing loss before receiving an implant, etiology of hearing loss (when known), and duration of implant use. The last two columns specify the stimulation mode and the implant type. Subjects gave informed consent and were paid on an hourly basis for their participation. All seven subjects had prior experience performing psychophysical tasks

TABLE I. Subjects. Subject identifying code, gender, age when tested for the present study, etiology of deafness (implanted ear), duration of bilateral severe-to-profound hearing loss prior to implantation, duration of implant use prior to the study, test electrode, stimulation mode, and device and electrode type. Electrodes are numbered in research order (rEL), which increases from apex to base. Nucleus implants have 22 electrode contacts and a maximum of 21 bipolar electrode channels; Clarion implants have 16 electrode contacts arranged in radial pairs (SPRL electrode) or in a straight line (HF electrode). Bipolar mode indicates the separation between active and reference electrodes in the stimulating pair: BP+1, BP+2, and BP+3 represent electrode separations of 1.5, 2.25, and 3.0 mm, respectively. For Clarion subjects, electrode types are SPRL=standard spiral electrode; HF=HiFocus electrode; HF+EPS=HiFocus electrode with electrode positioning system.

Subject code	M/F	Age	Etiology of deafness	Dur. (years)	Cl use (years)	Electrode (rEL)	Stimulation mode	Device (electrode)
N28	M	63	Meningitis	<1	6	12	BP+1	Nucleus 22
N30	F	63	Otosclerosis	10	6	17	BP+2	Nucleus 22
N31	M	82	Noise exposure; progressive SNHL	25	12	16	BP+3	Nucleus 22
N32	M	34	Maternal rubella; progressive SNHL	<1	5	12	BP+2	Nucleus 22
C06	M	65	Unknown; progressive SNHL	12	3	8	monopolar	Clarion C-I (SPRL)
C12	F	49	Otosclerosis	13	2	8	monopolar	Clarion C-I (SPRL)
C14	M	65	Unknown; progressive SNHL	47	2	8	monopolar	Clarion C-I (HF+EPS)
C16	F	48	Unknown; progressive SNHL	18	10 mo.	7	monopolar	Clarion C-I (HF)
C18	M	67	Otosclerosis	33	1.5	7	monopolar	Clarion C-I (HF+EPS)

and were given at least 2 h of practice before data collection commenced. All experiments in the study were approved by the Human Subjects Committee of the University of Minnesota Institutional Review Board.

2. Conditions and stimuli

Stimuli for Nucleus-22 users were generated by a PC connected through a parallel port to a BTNI cochlear implant interface (Shannon *et al.*, 1990). Stimulus amplitudes were specified in integer current step units (CSUs), which are uneven amplitude steps that vary between 0.07 and 0.30 dB for the range of current amplitudes used in the present experiment. CSUs were converted to calibrated current amplitudes using user-specific tables provided by Cochlear Corporation.

Stimuli for Clarion subjects were generated by a PC that controlled a special-purpose interface provided by Advanced Bionics Corporation for the Clarion C-I intracochlear stimulator. Stimulus amplitudes were specified in integer stimulus units (SUs), which are logarithmic amplitude steps of 0.1 to 0.3 dB for the range of amplitudes used here.[1] SU values were converted to calibrated amplitudes using a set of tables developed in our laboratory. This calibration compensates for nonlinearities in the current source that depend upon electrode impedance and pulse rate. Electrical impedances for Clarion subjects were measured at the beginning and end of each data collection session using the SCLIN for Windows clinical software running on a PC. Calibrated amplitudes for each test electrode were based on average impedance values across data collection sessions for a given experiment.

Increment-detection thresholds were measured using trains of biphasic pulses presented to an electrode in the middle or midbasal portion of the array. The initial phase of each pulse was always cathodic (negative). For the Nucleus-22 subjects, pulse duration was 80 μs/phase and pulse rate was 800 pulses/s. Stimuli were presented to Nucleus subjects in bipolar mode, using the electrode separations listed in Table I. The initial (cathodic) pulse was presented to the more basal electrode of the pair. For the Clarion users, pulse duration was 77 μs/phase and pulse rate was 1000 pulses/s. Stimuli were presented in monopolar mode, with an intracochlear electrode referenced to a ground electrode on the case of the receiver–stimulator. In the gated condition, the pulse train had a duration of 100 ms and its amplitude was incremented in the signal interval over its entire duration. In the fringe condition, the pedestal was 750 ms long and the increment to be detected occupied the final 100 ms. Increments were generated in the simulation software by specifying higher-amplitude pulses during the final 100 ms of pedestal duration. Thresholds for detecting increments were measured for five amplitude levels of the pedestal defined using a reference amplitude of 1 μA. The pedestal levels were evenly spaced in 20-log (μA) steps across the dynamic range of each listener's hearing. Thus, for all listeners, thresholds were measured at levels corresponding to approximately the same percentages of dynamic range expressed in dB.

3. Procedure

Each experimental session began by measuring the listener's absolute threshold and maximum acceptable loudness level (MAL) for the 100-ms pulse train used in the subsequent intensity-discrimination task. The threshold was measured using an adaptive 2-down, 1-up, three-interval forced-choice (3IFC) procedure that estimated the 70.7%-correct

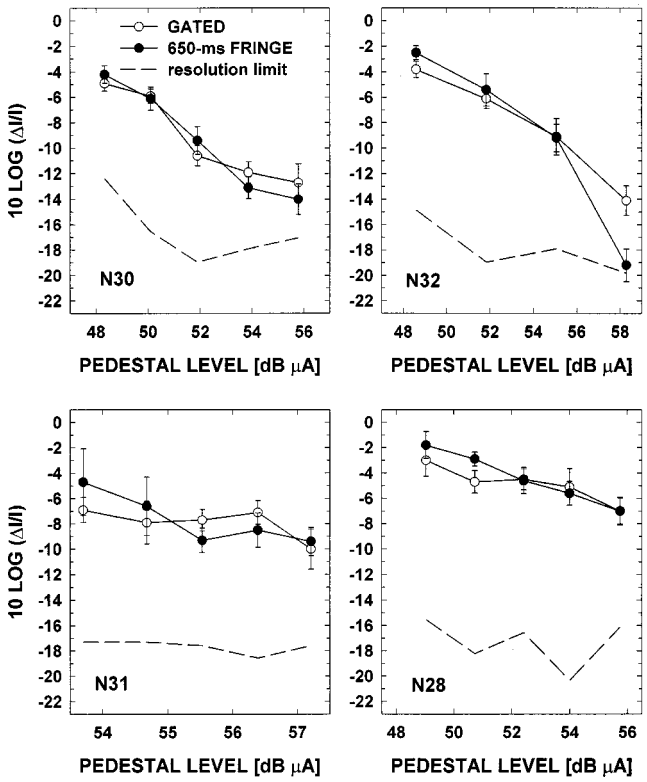


FIG. 2. Intensity DLs in Nucleus-22 users, measured in the gated (open circles) and 650-ms fringe (filled circles) conditions. The dashed line indicates the amplitude resolution limit of the cochlear implant.

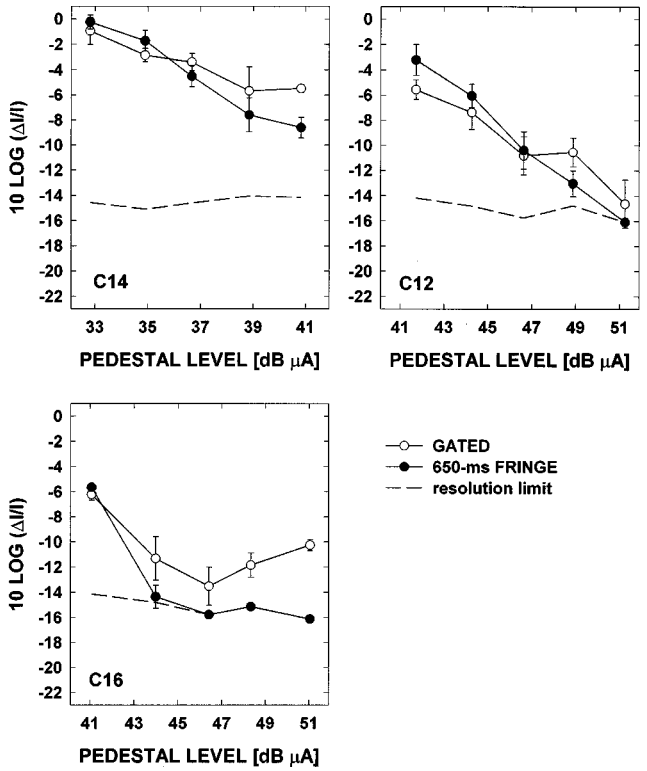


FIG. 3. Intensity DLs in Clarion users, measured in the gated (open circles) and 650-ms fringe (filled circles) conditions. The dashed line indicates the amplitude resolution limit of the cochlear implant.

point on the psychometric function (Levitt, 1971). The MAL was measured using an ascending method of adjustment. The difference between the level (in dB *re*: 1 μA) corresponding to the MAL and the level corresponding to the absolute threshold determined the listener's dynamic range for the tested stimulus. The estimated difference was used to compute five levels of the pulse train corresponding to 17%, 33%, 50%, 67%, and 83% of the dynamic range in dB, for which intensity discrimination was measured. Intensity DLs were measured using an adaptive 3IFC procedure similar to that used for absolute threshold. The observation intervals were separated by a 250-ms silent interval. On each trial, the increment (signal) occurred randomly in one of the three intervals. Initially the increment was large enough to be clearly detectable. After two consecutive correct responses the amplitude of the incremented portion of the pulse train was decreased by 2 units (CSUs for Nucleus users; SUs for Clarion users), and after an incorrect response it was increased by 1 unit. Visual feedback was provided after each response. The run was terminated after 12 reversals were obtained. Threshold was computed as the mean of the last eight reversals. Threshold amplitude increments were converted into Weber fractions in dB $(10 \log \Delta I/I)$, where I is proportional to the squared current amplitude (A^2) and represents electric power with a unit of $\mu A^2 \times$ ohm. The power increment ΔI is proportional to $2A\,\Delta A + (\Delta A)^2$, where ΔA represents the increment in current amplitude. The final threshold estimate was obtained by averaging four to six Weber fractions in dB obtained from separate runs.

4. Results

Figure 2 shows intensity DLs for individual Nucleus subjects. DLs are expressed in terms of the Weber fraction in dB and plotted as a function of pedestal level in dB *re:* 1μA. Error bars represent one standard deviation around the mean. Figure 3 shows the data obtained for individual Clarion subjects.

For all of the cochlear-implant listeners, Weber fractions tend to decrease with increasing pedestal level. This trend has been reported in the past studies of intensity discrimination in electric hearing (Shannon, 1983; Dillier *et al.*, 1983; Busby *et al.*, 1992; Nelson *et al.*, 1996; Donaldson and Viemeister, 2000). The present data show that a decrease in the Weber fraction with level is observed not only for gated stimuli (open symbols) but also when the increment occurs in an ongoing sound (filled symbols).

Apart from this general trend, slightly different patterns of results emerge for the two groups of cochlear-implant users. Overall, the Nucleus subjects (Fig. 2) do not appear to exhibit any systematic differences in sensitivity to intensity increments between the gated and fringe conditions. This observation is supported by statistical analysis, which showed no significant effect of condition across pedestal levels and subjects (ANOVA, $F = 0.41$, $p = 0.522$). Despite that general result, listener N32 exhibited a 5.1-dB smaller intensity DL for the fringe condition, and this difference may have been artificially limited by device resolution (see below in this section). The significance of this finding is unclear; it is possible that a similar pattern could be observed in other Nucleus users if a larger sample were tested. Note that for

subject N32, data were obtained at only four pedestal levels due to possible voltage compliance limitations at levels exceeding 60 dB re: 1 μA.

In contrast to the data from Nucleus users, the data from Clarion users do show some systematic differences between the intensity DLs measured with the gated pedestal and those measured with the preceding fringe (Fig. 3). The main effect of condition was significant (ANOVA, $F=36.61$, $p<0.001$). For listeners C14 and C12, the intensity DLs measured at two lowest levels with the 650-ms fringe are slightly higher than those measured with the gated pedestal. At the two highest levels, the intensity DLs for the 650-ms fringe fall below those for the gated pedestal. For subject C16, the intensity DLs measured in the fringe condition are consistently lower than those measured in the gated condition, except at the lowest pedestal level.

Because cochlear implants deliver current in quantized steps, they sometimes impose a limitation on users' intensity resolution. In Figs. 2 and 3, the dashed lines show Weber fractions that would be estimated if a subject consistently detected the smallest current increment that could be generated for a given pedestal amplitude. It is clear that device resolution sometimes prohibited assessment of the listeners' ''real'' intensity DLs in the fringe condition.[2] This occurred for subjects N32 and C12 at the highest pedestal level and for subject C16 at the three highest levels. Despite this limitation, it appears that for Clarion users, the 650-ms fringe causes a reduction in intensity DL (and thus the Weber fraction) at higher pedestal levels.

It is not clear why a similar difference in the intensity DLs measured with and without the preceding fringe was not observed in the Nucleus users. Small differences in the parameters of pulse trains used for Nucleus versus Clarion users are unlikely to explain the discrepancy. Also, the fact that the Clarion users were generally stimulated with lower current amplitudes does not seem to provide an explanation, since there is some overlap of current amplitudes tested in Nucleus and Clarion subjects and, at similar levels, different results are still observed (e.g., compare Nucleus user N30 for levels around 48–51 dB *re*: 1 μA and Clarion user C16 for the same range of levels). The most obvious difference was in the mode of stimulus presentation (medial monopolar for Clarion subjects versus bipolar for Nucleus subjects). We do not know whether differences in electrode coupling can account for the observed discrepancy between the two groups. We will return to this issue in the Discussion.

Overall, the present data indicate that at medium and high stimulus levels the Weber fractions reported in previous cochlear-implant studies may underestimate intensity resolution for ongoing sounds. In particular, our data show that when a pedestal presented at medium to high levels has been on for 650 ms before an increment occurs, intensity DLs may be reduced significantly relative to the gated case. Assuming that neural adaptation underlies the improvement in intensity DLs when a preceding fringe is added, this result suggests that adaptation central to the inner hair cell synapse may play a role in the gated-continuous difference observed in normal hearing.

B. Gated versus fringe conditions in normal hearing

As mentioned in the Introduction, it is believed that in normal hearing, the difference between the intensity DL measured with the gated pedestal and the intensity DL measured with the continuous pedestal stems from neural adaptation. It is not clear how much time is needed from the onset of the pedestal for the adaptation to cause a significant decrease in the intensity DL. Most studies of increment detection in normal-hearing listeners used pedestals that were played continuously throughout the experimental run and compared the observed thresholds with those for pedestals gated with the increments. Scharf *et al.* (1992) studied the effect of the duration of a preceding fringe on increment detection, but they did not use fringe durations shorter than 1 s. Moore and Peters (1997) used shorter fringe durations (10 and 200 ms) but they did not measure intensity DLs using the gated (no-fringe) paradigm. The current experiment measured intensity DLs in normal-hearing subjects as a function of fringe duration.

1. Subjects

Three listeners with normal hearing participated in the study. Their absolute thresholds were within 10 dB of laboratory norms at octave frequencies between 250 and 8000 Hz. One listener (S1) was the first author, and the two other listeners were paid for their services. The listeners had previous experience in various psychoacoustic tasks.

2. Conditions and stimuli

Intensity DLs were measured for a gated condition, a continuous condition, and two fringe conditions. In the fringe conditions, the pedestal onset preceded the increment onset by 250 ms in one case and 650 ms in another. The increment was always 100 ms in duration. The pedestals and increments were gated with 5-ms raised-cosine ramps. Thresholds were measured for two types of stimuli, a 1-kHz tone and a noise that was low-pass filtered at 5 kHz with a 6-dB/oct attenuation outside the passband. The tone was presented at a level of 58 dB SPL and the noise was presented at a spectrum level of 20 dB SPL measured at 1 kHz.

For all conditions except the continuous-noise condition, the stimuli were generated digitally on a NeXT computer using a 16-bit D/A converter and a sampling rate of 44.1 kHz. The intensity increment in a tonal pedestal was produced by in-phase addition of a 100-ms 1-kHz tone to the pedestal. For the noise pedestal, the increment was obtained by adding an independent sample of the noise to the pedestal. The continuous noise pedestal was produced by an analog noise generator (General Radio 1381). The increment was obtained by mixing the continuous pedestal with a computer-generated 100-ms burst of noise. Fixed analog attenuators were used to achieve the desired levels of presentation. Stimuli were presented monaurally through Sony MDR-V6 headphones.

3. Procedure

An adaptive 3IFC 2-down, 1-up procedure was used to measure intensity DLs in the three normal-hearing subjects.

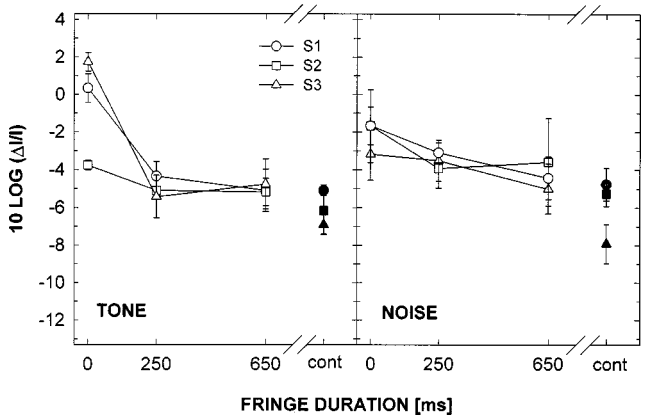


FIG. 4. Individual thresholds for intensity discrimination for three listeners with normal hearing, plotted as a function of the duration of a fringe preceding the increment. The left panel shows data for a 1-kHz tone, and the right panel shows data for a wideband noise, low-pass filtered at 5 kHz. The filled unconnected symbols represent intensity DLs measured with a continuous pedestal.

For the tonal pedestal, the increment was adjusted in steps of 4 dB $[10\log(\Delta I/I)]$ until four reversals were obtained. The step size was then reduced to 2 dB for the subsequent eight reversals. The same stepping rule was applied for the noise, except that in this case the level of the added sample of noise rather than the Weber fraction was varied adaptively. Visual feedback indicating the correct interval was provided after each trial. The block terminated after 12 reversals were completed. The threshold was computed as the mean of the last eight reversals. The final threshold estimate was obtained by averaging thresholds from three single runs. All conditions were first run using the tonal pedestal and then the same conditions were repeated using the noise pedestal. The order in which the experimental conditions were tested for a given type of stimulus was different for different listeners.

4. Results

Figure 4 shows individual data for the normal-hearing listeners. Weber fractions are plotted as a function of fringe duration, where 0-ms fringe corresponds to the gated condition. Data for the gated and fringe conditions are represented by open symbols; data obtained for the continuous pedestal are represented by the filled symbols at the right of each panel.

For the tonal pedestal (left panel), the decrease in intensity DL was largest between the gated case and the 250-ms fringe condition. Increasing the fringe duration from 250 to 650 ms did not appear to further affect the detection of the increment. The average difference between the intensity DL measured with the gated pedestal and the intensity DL measured with the 650-ms fringe averaged across the listeners was 4.4 dB ($F=587$, $p<0.001$), which is very close to the 4.6-dB gated-continuous difference reported by Viemeister and Bacon (1988). The average decrease in threshold between the 650-ms fringe condition and the continuous condition was only 1.1 dB and was not statistically significant ($F=11.05$, $p=0.006$). In contrast to our result, Scharf *et al.* (1992) found that for a 4-kHz tone, intensity DLs continued to decrease as fringe duration increased up to 30 s. In Fig. 4, the intensity DL measured with a 650-ms preceding fringe is already very similar to that measured with a continuous pedestal. We have no explanation for the discrepancy between the present results and their data. Scharf *et al.* used a higher frequency (4 kHz) than that used in the present study (1 kHz). However, as shown by Bacon and Viemeister (1994), the frequency dependence of adaptation may have no effect on the way the adapting stimulus behaves as a masker. Bacon and Viemeister compared intensity DLs measured in the same subjects using a 1-kHz tone and a 16-kHz tone for gated and continuous pedestals, respectively. They found that despite very strong loudness adaptation for the 16-kHz tone (that decayed to inaudibility), similar differences between gated and continuous thresholds were observed for the two pedestal frequencies.

The data shown in the right panel of Fig. 4 were obtained for the noise pedestal. A decrease of 2.1 dB in average intensity DL was observed between the gated condition and the 650-ms fringe condition, but this decrease was not statistically significant ($F=2.88$, $p=0.083$). However, the gated intensity DLs were significantly different from those obtained with a continuous pedestal ($F=40$, $p<0.001$). The average gated-continuous difference for the low-pass-filtered noise was 4.2 dB. Thus, it appears that with increasing fringe duration, intensity DLs decrease at a slower rate for noise stimuli than for tones. Using noise pedestals, Scharf *et al.* found essentially no change in intensity DLs with increasing fringe duration up to 30 s. Assuming that adaptation is the primary mechanism underlying the change in sensitivity to intensity increments between the gated and continuous conditions, it appears that for any given fringe duration, less adaptation is produced by noise than by tonal stimuli.

For listeners with normal hearing, the Weber fractions were measured only at one selected pedestal level that was slightly below the level corresponding to the middle of these listeners' dynamic range. Generally, the observed Weber fractions were comparable or worse than the Weber fractions observed for cochlear-implant users for midrange pedestal levels. However, a direct comparison is difficult since it cannot be made at equivalent levels of pedestal presentation, and it is not clear to what extent the acoustic stimuli can be considered equivalent to the pulse-train stimuli used with implant users.

III. EXPERIMENT 2

In this experiment, the gated and continuous paradigms were used to test five cochlear-implant users in the intensity-discrimination and increment-detection tasks. This experiment further investigated the role of long-term stimulation in improving sensitivity to intensity increments. In particular, it tested whether a continuous pedestal would produce gated-continuous differences similar to those observed in normal hearing.

A. Subjects

Five users of the Clarion cochlear implant participated in this experiment. Three of them served as subjects in experiment 1. More detailed information about the listeners is given in Table I. The listeners had earlier experience in psy-

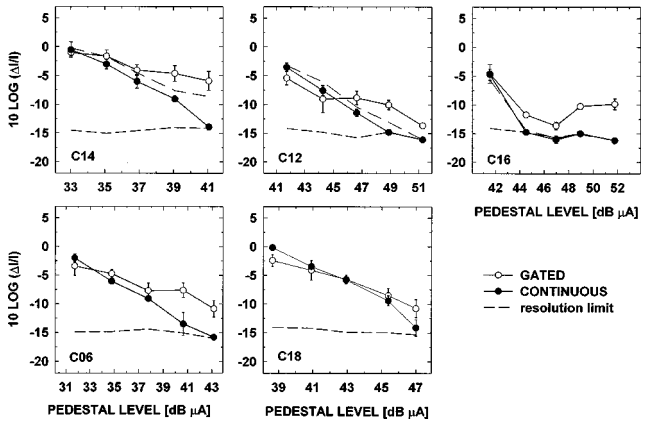


FIG. 5. Intensity DLs measured in Clarion users for a pedestal gated with the increment (open symbols) and a pedestal played continuously throughout the run (filled symbols). The dashed-dotted line in the upper row represents thresholds for the 650-ms fringe, replotted from Fig. 3. The dashed line shows the resolution limit of the device.

chophysical tasks and were given at least 2 h of practice in the intensity-discrimination and increment-detection tasks. Subjects gave informed consent and were compensated for their services.

B. Stimuli and procedure

The stimuli and the procedure used in this experiment were identical to those used for the Clarion users in experiment 1. The only difference was that instead of using a fixed-duration fringe preceding the intensity increment, the pedestal was present continuously throughout the run. Three observation intervals and the correct-response interval were signaled visually on each trial. A 100-ms increment in pulse-train intensity appeared randomly in one of the observation intervals. Although intensity DLs for the gated pedestal were previously measured in experiment 1 for three of the listeners, new data were collected from them for this condition along with the data for the continuous pedestal. This was done to ensure that any differences between the intensity DLs obtained using the two experimental paradigms did not result from changes in the listeners' performance across sessions. In this experiment, only three separate threshold estimates were averaged to obtain the final threshold.

C. Results

Figure 5 shows data for the five users of the Clarion cochlear implant. At the lowest level, intensity DLs measured with the continuous pedestal (filled symbols) are similar to or fall slightly above the DLs measured with the gated pedestal (open symbols). As the pedestal level increases, the intensity DLs generally decrease. An exception occurs for listener C16, whose thresholds decrease for low and medium levels, and then increase slightly at high levels.

For all of the listeners, continuous thresholds decrease at a faster rate with increasing level than gated thresholds. As a consequence, at high levels, intensity DLs measured with the continuous pedestal are lower (often substantially) than intensity DLs measured with the gated pedestal. Thus, at high levels all five of the Clarion users exhibit a clear gated-continuous difference. The maximum size of the gated-continuous difference could not be estimated because intensity DLs for the continuous pedestal were limited by amplitude resolution of the cochlear implant. Even with this limitation, however, the gated-continuous difference was as large as 8 dB for C14 and 6.4 dB for C16. In contrast to the normal-hearing data of Viemeister and Bacon (1988), the present data appear to indicate that in electric hearing, the gated-continuous difference in Weber fractions is not constant across a range of suprathreshold pedestal levels but continues to increase over the entire dynamic range. All of the listeners reported that the continuous pedestal decayed in loudness during the experimental run, frequently to inaudibility. This strongly suggests an involvement of long-term adaptation. It should be noted that this adaptation could not originate from the depletion of neurotransmitter at the inner hair cell synaptic site, since hair cells are by-passed in electric hearing. Instead, the data suggest that a strong adaptation of more central origin affects increment detection in cochlear-implant listeners.

Since three of the listeners (C14, C12, and C16) participating in this experiment also completed experiment 1, their data can be compared for the 650-ms fringe condition (the dashed-dotted line in Fig. 5) and the continuous condition (filled symbols in Fig. 5). This comparison reveals a similar pattern of differences between the Weber functions obtained for the gated condition and the functions obtained for the fringe and continuous conditions. At higher levels the Weber fractions observed for the continuous pedestal are slightly lower than those measured with the fringe (when the latter were not limited by the amplitude resolution of the implant). The Weber fractions observed in the gated condition were highly repeatable (the average differences in dB between the sets of gated thresholds shown in Figs. 3 and 5 are 0.68, 1.06, and 0.82, for C14, C12, and C16, respectively). This indicates that the observed decrease in the intensity DL (continuous versus fringe condition) was unlikely the result of training effects.

IV. DISCUSSION

In cochlear-implant users, intensity discrimination expressed in terms of the Weber fraction improves with increasing stimulus level, similar to what is observed for listeners with acoustic hearing. In normal hearing, a decrease in the Weber fraction with increasing pedestal level is thought to reflect nonlinear spread of excitation on the basilar membrane (see Viemeister, 1988). Intensity-discrimination data from cochlear-implant users suggest nonlinear spread of excitation across the neural population despite the lack of basilar-membrane processing. A possible explanation for the nonlinear spread observed in electric hearing was offered by Nelson *et al.* (1996). They suggested that at low levels of stimulation, only sparsely populated neurons close to the electrode respond. Assuming that some criterion increase in the total spike count is necessary for increment detection, a relatively large change in stimulus level is required to achieve the criterion increase. At moderate and high levels, the current spreads into the modiolus and activates the spiral nerve bundle located farther away from the electrode. Con-

sequently, the rate-intensity functions become steeper and a smaller change in stimulus intensity is sufficient to reach the criterion increase in total spike count.

Experiment 1 revealed that at high stimulus levels, a 650-ms fringe preceding an increment in stimulus intensity improves increment detection in users of the Clarion cochlear implant but does not affect increment detection in users of the Nucleus-22 implant. As noted earlier, the mode of stimulation was different for the two groups of implant users. Monopolar stimulation, used by Clarion subjects in this study, is known to produce a broader excitation pattern than bipolar stimulation, used by the Nucleus-22 subjects. This has been demonstrated physiologically at peripheral (Brown *et al.*, 1996; Kral *et al.*, 1998) and central sites including the auditory cortex (Bierer and Middlebrooks, 2002). However, monopolar stimulation may result in lower average response probabilities for auditory-nerve fibers (Pfingst *et al.*, 1997; Miller *et al.*, 2002). Intuitively, lower response probabilities would be more likely to reduce rather than increase adaptation, and thus lead to less facilitation of increment detection in the Clarion users. Our data showed the opposite effect, namely a larger gated-fringe difference in Clarion users than in Nucleus users. A notion that broader patterns of excitation produce more adaptation at central sites would be consistent with our findings, but we are not aware of any documented evidence that would support this notion.

The results of experiment 2 suggest that strong long-term adaptation occurs in cochlear-implant users despite the lack of peripheral processing involving inner hair cell synaptic transmission. Strong loudness decay of the continuous stimulus, often to inaudibility, supports this conjecture. Even though no direct link between loudness decay and the size of the gated-continuous difference has been identified, both effects are believed to reflect adaptation (Scharf *et al.*, 1992; Bacon and Viemeister, 1994). The slightly smaller difference between intensity DLs obtained in the gated versus fringe conditions (Fig. 3) compared with that difference for the gated versus continuous conditions (Fig. 5) may suggest that adaptation is not complete after 650 ms of stimulus presentation. Also, none of the subjects reported loudness decay in the task using the 650-ms fringe.

A. Predictions based on subtractive adaptation (constant ΔI)

As mentioned before, it is not clear exactly how adaptation would aid increment detection. Bacon and Viemeister (1994) considered two models of adaptation. One model assumed subtractive adaptation. According to this model, the excitation produced by the masker/pedestal would be subtracted due to adaptation and, independent of masker level, would fall to some fixed excitation level below absolute threshold. Since a long exposure to the pedestal frequently renders it inaudible, the task becomes effectively a detection-in-quiet task. Thus, the model of subtractive adaptation predicts that a constant level of the signal (added to the pedestal in order to produce an increment) is required to reach threshold. Bacon and Viemeister plotted their threshold increments in terms of the level of the ''added'' signal and found that subtractive adaptation did not predict their data correctly.

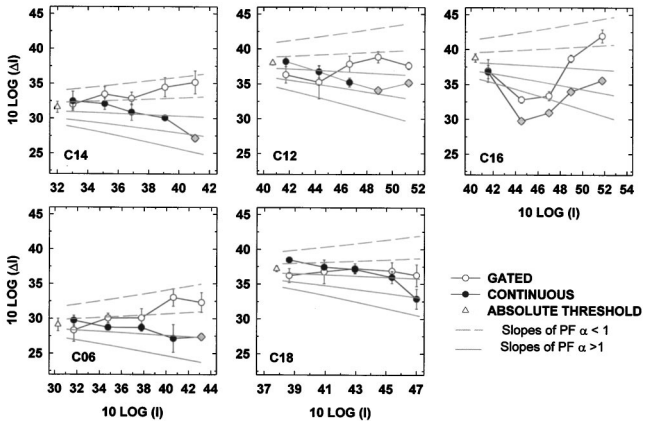


FIG. 6. Thresholds from Fig. 5 replotted in terms of 10 log (ΔI) as a function of pedestal level for the gated (open circles) and the continuous (filled circles) pedestal. The leftmost unconnected open symbols represent thresholds for absolute detection of the pulse train. Thresholds limited by the resolution of the cochlear implant are represented by gray diamonds. Gray lines represent predictions based on Eq. (1). The predictions were obtained for exponents α equal 0.7, 0.9, 1.1, 1.3, and 1.5, where exponent 0.7 is represented by the upper dashed line, and exponent 1.5 is represented by the lowest solid line. For listener C06 the line representing the exponent of 1.5 is not included.

Rather than being constant, the signal level necessary to produce a threshold increment increased in proportion to the pedestal level.

Therefore, they considered a different model that was based on attenuative adaptation. The assumption was that the response to the pedestal and the response to the pedestal-plus-increment are both attenuated by the same factor that is inversely proportional to the pedestal level. In this case, the level of the added signal must increase in proportion to the pedestal level in order to produce excitation necessary for absolute detection of that signal. This approach accurately described the behavior observed in Bacon and Viemeister's data. Bacon and Viemeister demonstrated that the result predicted by attenuative adaptation could also be predicted by subtractive adaptation preceded by a logarithmic nonlinearity. A compressive nonlinearity presumably is not operating at the stages of auditory processing that are central to the cochlea, and thus it is not present in the processing by the auditory system of cochlear-implant users. If the model of subtractive adaptation preceded by a nonlinearity is correct, then subtractive adaptation should predict the results obtained in electric hearing, i.e., the level of the signal that must be added to the pedestal to produce a just-detectable increment should be constant across the listener's dynamic range. Figure 6 shows the intensity-discrimination thresholds replotted from Fig. 5 and expressed in terms of the power increment in dB ($10 \log \Delta I$). The thresholds in Fig. 6 can also be thought of as the levels of the signal orthogonal to the pedestal that produce just-noticeable increments when the signal is added to the pedestal.

As in Fig. 5, the filled symbols represent the data for the continuous pedestal and the open symbols represent the gated condition. The leftmost unconnected symbols show thresholds for detecting the pulse train in quiet. It appears that none of the listeners required a constant level of the added signal to produce the threshold increment in the con-

tinuous condition. As the pedestal level increased, a relatively small albeit consistent decrease in threshold increment was observed except for the cases where the intensity DL was limited by the resolution of the device. This result does not support the idea of subtractive adaptation. The data are also inconsistent with the idea of attenuative adaptation as described by Bacon and Viemeister. Some listeners show an increase in threshold level that may be roughly proportional to the pedestal level (C14, C12, C06, and C16 at mid to high levels) but only for the gated condition, for which long-term adaptation does not occur or at best is negligible. The increase in threshold $\Delta A\,[\mathrm{dB}\,re{:}1\ \mu\mathrm{A}]$ (and thus also ΔI) with increasing level of the pedestal was also observed in other studies that measured intensity discrimination as a function of level using gated pedestals (Shannon, 1983; Nelson *et al.*, 1996).

B. Explanation in terms of Laming's theory

Similar to the acoustic data (e.g., Hanna *et al.*, 1986; Viemeister and Bacon, 1988; Bacon and Viemeister, 1994), there is a trend for intensity DLs measured in electric hearing to be larger in the continuous than in the gated condition at low pedestal levels, but the trend is reversed at higher pedestal levels. As shown by Hanna *et al.*, this result can be predicted by Laming's theory (Laming, 1986), at least qualitatively. The theory, originally developed to explain the phenomenon of negative masking (discussed in the following section), assumes that a nonlinear (square-law) transformation combined with differential coupling operates on near-threshold stimuli. Related to this is the assumption that a system responds to changes in amplitude/intensity rather than the absolute amplitude of a signal, and thus it is compatible with the idea of a change detector (MacMillan, 1971, 1973; Hafter *et al.*, 1997). Laming's approach correctly predicts steeper psychometric functions and larger intensity DLs observed for the continuous versus gated pedestals near absolute threshold. As explained by Hanna *et al.* (1986), applying Laming's assumptions to the gated condition would lead to the mean value for the decision variable that would be proportional to $(I+\Delta I)^2$ in the signal interval, and the mean value proportional to I^2 in the nonsignal interval. Consequently, under the assumption of a constant variance, d' would be proportional to the difference between the two variables, which is $2I\Delta I+(\Delta I)^2$. In the continuous (or, more generally, nongated) condition, the mean value of the decision variable would be proportional to $(\Delta I)^2$ in the signal interval and to zero in the nonsignal interval, resulting in d' being proportional to just $(\Delta I)^2$. Any value of $(\Delta I)^2$ would lead to a higher d' (and thus, to a lower threshold) in the gated versus nongated condition. This is generally observed in the data for the lowest pedestal level shown in Fig. 5 (the exception is subject C16).

At higher pedestal levels, the square-law transformation would be applied only in the nongated condition, where ΔI is always near threshold. In the gated condition, each onset is considered a change in intensity. Since these changes are suprathreshold, the square-law nonlinearity does not apply. As a result, lower thresholds (higher d' for a given ΔI) should be observed in the nongated versus gated condition. This trend is observed in the data at higher pedestal levels.

C. Negative masking

For the continuous pedestal, the level of the signal in Fig. 6 decreases below the threshold for detection of that signal in quiet. This phenomenon has been observed in acoustic hearing for tonal stimuli and has been dubbed ''negative masking'' because the presence of a simultaneous masker appears to lower the threshold for detecting the signal instead of raising it (Hanna *et al.*, 1986; Viemeister and Bacon, 1988; Bacon and Viemeister, 1994). There are, however, some differences between the negative masking in acoustic hearing and the negative masking observed in electric hearing. In acoustic hearing, this phenomenon has been observed exclusively for the gated thresholds and at masker levels near absolute threshold. At higher masker levels, thresholds for detecting the signal are elevated by the presence of the masker in both the gated and continuous condition.

In the listeners with cochlear implants, the signal levels necessary to reach intensity DLs at low pedestal levels for the two conditions fall around the threshold measured in quiet, with the tendency for the threshold to be higher for the continuous pedestal than for the gated pedestal. As the masker level increases, the signal levels at threshold for the gated masker stay near or increase above the absolute threshold. In contrast, the signal levels measured with the continuous masker fall progressively further below absolute threshold as masker level increases. In some cases there is an upturn in the data for the continuous condition at high levels (C12, C06, C16), but it almost certainly results from the limited amplitude resolution of the cochlear implant (see Fig. 5). Thus, at high levels no negative masking is observed for the gated masker and an increasing amount of negative masking is observed for the continuous masker. This result is *not* predicted by Laming's theory, which does not predict any negative masking for the continuous condition.

D. Explanation based on psychometric functions for detection

An interesting possible account for the observed data is in terms of the underlying psychometric functions for detection. Assume, as suggested by Tanner (1956), that the discriminability of an intensity change is equal to the difference between the detectabilities of the incremented pedestal and the pedestal alone

$$d'_{\mathrm{disc}}(\Delta I)=d'_{\mathrm{det}}(I+\Delta I)-d'_{\mathrm{det}}(I). \tag{1}$$

The psychometric function for detection both in acoustic hearing (Egan *et al.*, 1969) and in electric hearing (Donaldson *et al.*, 1997; Donaldson and Viemeister, 2001) can be well described by $d'_{\mathrm{det}}(I)=kI^{\alpha}$, where α is the slope of the psychometric function on log–log coordinates. These equations lead to

$$d'_{\mathrm{disc}}(\Delta I)=k[(I+\Delta I)^{\alpha}-(I)^{\alpha}]. \tag{2}$$

Predictions based on Eq. (2) are shown in Fig. 6. Dashed lines represent predicted thresholds for α values of 0.5 and

0.7, and solid lines represent predicted thresholds for α values of 1.1, 1.3, and 1.5. The predictions were obtained by solving Eq. (2) for ΔI and adjusting d'_{disc}/k so that ΔI equals the observed detection threshold when $I=0$. As demonstrated in Fig. 6, for $\alpha>1$, Eq. (2) predicts increasing negative masking as the pedestal level increases. For $0<\alpha<1$, it predicts an increasing amount of masking with increasing level. Although this appears promising, the values for α that could provide at least a rough account of the data appear to disagree with the data on detection. Donaldson and Viemeister (2001) obtained detailed psychometric functions for detection of 300-ms, 100-pulses/s trains in 10 adults with Nucleus-2 implants. Generally, the slopes of the psychometric functions decreased with increasing pulse width. For 80-μs/phase pulses, close to the values used in the present study, the estimates of α ranged from approximately 1.2 to 8, with a mean of 3. These slopes, which are much larger than those for acoustic hearing, predict considerably more negative masking than shown by the data. Furthermore, since these predictions are based on psychometric functions for gated stimuli they technically are appropriate only for the gated pedestals. In that condition, the data, except for those of C16, show little or no negative masking.

It is not clear how to explain the discrepancy between the predictions of Eq. (2) and the data. The most likely possibility is that Eq. (1) is incorrect. Perhaps because of the limited range over which d'_{det} can be measured, the validity of Eq. (1) has not been empirically assessed either in acoustic or electric hearing. The observation that intensity discrimination cannot be predicted from the psychometric function for absolute detection may be tapping something fundamental about the mechanisms involved in the two tasks. It may suggest that intensity discrimination is based on change detection, whereas absolute detection is mediated by a different mechanism. Far less likely is the possibility that the differences in implant type, pulse rate, stimulus duration, and individual differences account for the discrepancy between the predictions and the slopes resulting from the data of Donaldson and Viemeister.

E. Adaptation in response to auditory and electric stimulation

The size of the gated-continuous difference in the upper part of the dynamic range of hearing is larger for Clarion users than it is for normal-hearing listeners. This may reflect stronger adaptation due to electrical stimulation of the nerve. Kiang and Moxon (1972) demonstrated that neurons are driven to respond at higher rates with electrical stimulation than with auditory stimulation. Shannon (1983) observed strong loudness decay in cochlear-implant users for sinusoids with frequencies higher than 300 Hz. He suggested that stronger adaptation occurs when neurons are driven electrically because they are driven to higher rates and have less time for recovery. In contrast, neurons driven with auditory stimulation respond stochastically at lower rates and have more time to recover. Consequently, auditory stimulation causes less adaptation. Shannon's observations were not, however, supported by the study of Brimacombe and Eisenberg (1984), who generally did not observe loudness decay in cochlear-implant listeners stimulated with high-frequency sinusoids (even as high as 16 kHz). Only three of their 17 subjects showed loudness decay. Those three subjects lost their hearing at earlier ages and did not use hearing aids for a longer time than subjects who could sustain the perception of a tone for at least 2 min. For this reason, Brimacombe and Eisenberg concluded that neural degeneration was a probable cause of loudness decay. The Clarion users participating in the present study reported a considerable decay in stimulus loudness during each block of trials, often to inaudibility, but there was no apparent relationship between the degree of loudness decay and years of deafness. As in normal-hearing listeners, no direct link between loudness decay and the size of the gated-continuous difference was found in the Clarion users. It may appear as if two different mechanisms underlie the two phenomena. More likely, however, loudness adaptation and the gated-continuous difference are a result of different types of adaptation that occur at different levels of processing and are characterized by different time constants. Loudness decay may be a result of a long-term adaptation occurring at central sites, whereas the gated-continuous difference may reflect short-term adaptation determined more peripherally.

V. FINAL REMARKS AND CONCLUSIONS

The gated-continuous difference observed for tones in listeners with normal hearing does not change significantly for fringe durations longer than 650 ms. In fact, the gated-continuous difference and the difference measured with a 250-ms fringe are already very similar. In listeners with electric hearing, a larger difference in intensity DLs is observed for the gated versus continuous condition than for the gated versus 650-ms fringe condition. This may be due to the difference in the type of stimulus. This possibility is supported by the noise data collected from normal-hearing subjects, which show no significant decrease in the Weber fraction between the gated and the 650-ms fringe condition, and a significant decrease between the 650-ms fringe and the continuous condition. For noise stimuli, excitation is distributed across neurons with different CFs. Broader distribution of excitation and lower level of stimulation at specific CFs may produce less adaptation and a slower temporal course of adaptation compared with that for tones. In cochlear-implant listeners, neurons were driven at higher rates by the electric pulse trains due to increased synchrony and, at least with monopolar stimulation, a broader distribution of excitation across neurons was likely produced compared with that produced by acoustic stimulation by tones. These differences could contribute to the more sizable gated-continuous difference in the Clarion users.

It is likely that in acoustic hearing, peripheral synaptic adaptation and adaptation occurring at more central sites of processing both play a role in facilitating detection of intensity changes in an ongoing stimulus. It may be that the time course of synaptic adaptation is faster than that for central adaptation, and that synaptic adaptation becomes complete at shorter fringe durations.

In summary, the following conclusions can be drawn from our experiments:

(1) Cochlear-implant users exhibit an increased sensitivity to intensity increments after a prolonged exposure to the stimulus. Thus, DLs reported in other studies that measured intensity discrimination in cochlear-implant users with gated pedestals may underestimate these listeners' ability to detect small changes in current amplitude of an ongoing stimulus. This may be true particularly when the exposure times exceed 650 ms and the level of stimulus presentation is relatively high in the listener's dynamic range.
(2) The gated-continuous difference in sensitivity to intensity increments depends on the level in the listener's dynamic range. At high levels, the listeners' ability to detect small changes in the current amplitude after prolonged exposure to a stimulus may be limited by the amplitude resolution of the cochlear implant.
(3) Strong long-term adaptation is observed in cochlear-implant users. It reveals itself through strong loudness decay and through an increased sensitivity to intensity increments. This adaptation, combined with a change-detection mechanism, may be the primary cause of the gated-continuous difference in intensity-discrimination thresholds.
(4) Adaptation originating at sites central to the synapse between the IHCs and the spiral ganglion is likely to be a strong contributing factor in creating the gated-continuous difference in normal hearing.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC00110 and the Lions 5M International Hearing Foundation. John Van Essen converted Robert Shannon's computer software into the C language and modified that software for testing Nucleus subjects. Eric Javel developed the experimental software for Clarion subjects which made use of a research interface provided by Advanced Bionics Corporation. Cochlear Corporation provided calibration tables for Nucleus subjects. Heather Kreft, Suzanne Hansel, Kim Fishman, and Shanna Allen assisted with data collection. The authors extend special thanks to the cochlear-implant and normal-hearing listeners who participated in these experiments. We also thank Bob Shannon and an anonymous reviewer for their helpful comments on an earlier version of this paper.

[1]Nominal amplitude steps for the Clarion C-l device are a constant 0.3 dB across amplitude. However, the actual step size decreases with increasing electrode impedance and at high current amplitudes where the current source begins to saturate. At current levels used in the present study, the current source was essentially linear.

[2]During the track when the listener correctly detected the increment on three consecutive trials at the resolution limits of the device, the program assumed an incorrect response, counted it as a reversal, and continued presentations at the same level. If the subject always responded correctly when this smallest current increment was presented, then his discrimination threshold was computed to be half of the size of this resolution step.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Simultaneous masking by gated and continuous sinusoidal maskers,'' J. Acoust. Soc. Am. **78**, 1220–1230.
Bacon, S. P., and Viemeister, N. F. (**1994**). ''Intensity discrimination and increment detection at 16 kHz,'' J. Acoust. Soc. Am. **95**, 2616–2621.
Bierer, J. A., and Middlebrooks, J. C. (**2002**). ''Auditory cortical images of cochlear-implant stimuli: Dependence on electrode configuration,'' J. Neurophysiol. **87**, 478–492.
Brimacombe, J. A., and Eisenberg, L. S. (**1984**). ''Tone decay in subjects with the single-channel cochlear implant,'' Audiology **23**, 321–332.
Brown, C. J., Abbas, P. J., Borland, J., and Bertschy, M. R. (**1996**). ''Electrically evoked whole nerve action potentials in Ineraid cochlear implant users: Responses to different stimulating electrode configurations and comparison to psychophysical responses,'' J. Speech Hear. Res. **39**, 453–467.
Busby, P. A., Tong, Y. C., and Clark, G. M. (**1992**). ''Psychophysical studies using a multiple electrode cochlear implant in patients who were deafened early in life,'' Audiology **31**, 95–111.
Campbell, R. A., and Lasky, E. Z. (**1967**). ''Masker level and sinusoidal-signal detection,'' J. Acoust. Soc. Am. **42**, 972–976.
Dillier, N., Spillmann, T., and Guntensperger, J. (**1983**). ''Computerized testing of signal encoding strategies with round window implants,'' in *Cochlear Prostheses: An International Symposium*, edited by C. W. Parkins and S. W. Anderson (Ann. N.Y. Acad. Sci., New York), Vol. 405, pp. 360–369.
Donaldson, G. S., and Viemeister, N. F. (**2000**). ''Intensity discrimination and detection of amplitude modulation in electric hearing,'' J. Acoust. Soc. Am. **108**, 760–763.
Donaldson, G. S., and Viemeister, N. F. (**2001**). ''Effects of pulse duration on psychometric functions for detection and temporal integration functions in cochlear-implant listeners,'' Conference on Implantable Auditory Prostheses, Pacific Grove, CA.
Donaldson, G. S., Viemeister, N. F., and Nelson, D. A. (**1997**). ''Psychometric functions and temporal integration in electric hearing,'' J. Acoust. Soc. Am. **101**, 3706–3721.
Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). ''Masking-level differences and the form of the psychometric function,'' Percept. Psychophys. **6**, 209–215.
Geisler, C. D., Le, S., and Schwid, H. (**1979**). ''Further studies on the Schroeder–Hall hair-cell model,'' J. Acoust. Soc. Am. **65**, 985–990.
Green, D. M. (**1969**). ''Masking with continuous and pulsed sinusoids,'' J. Acoust. Soc. Am. **46**, 939–946.
Green, D. M., Nachmias, J., Kearney, J. K., and Jeffress, L. A. (**1979**). ''Intensity discrimination with gated and continuous sinusoids,'' J. Acoust. Soc. Am. **66**, 1051–1056.
Hafter, E. R., Bonnel, A.-M., and Gallun, E. (**1997**). ''A role for memory in divided attention between two independent stimuli,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, and A. Q. Summerfield (Whurr, London), pp. 199–208.
Hanna, T. E., von Gierke, S. M., and Green, D. M. (**1986**). ''Detection and intensity discrimination of a sinusoid,'' J. Acoust. Soc. Am. **80**, 1335–1340.
Hochmair-Desoyer, I. J., Hochmair, E. S., Burian, K., and Fisher, R. E. (**1981**). ''Four years of experience with cochlear prostheses,'' Med. Prog. Technol. **8**, 107–119.
Javel, E. (**1996**). ''Long-term adaptation in cat auditory-nerve fiber responses,'' J. Acoust. Soc. Am. **99**, 1040–1052.
Kiang, N. Y.-S., and Moxon, E. C. (**1972**). ''Physiological considerations in artificial stimulation of the inner ear,'' Ann. Otol. Rhinol., Laryngol. **81**, 714.
Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (**1998**). ''Spatial resolution of cochlear implants: The electrical field and excitation of auditory afferents,'' Hear. Res. **121**, 11–28.
Laming, D. (**1986**). *Sensory Analysis* (Academic, London).
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Macmillan, N. A. (**1971**). ''Detection and recognition of increments and decrements in auditory intensity,'' Percept. Psychophys. **10**, 233–238.
Macmillan, N. A. (**1973**). ''Detection and recognition of intensity changes in tone and noise: The detection-recognition disparity,'' Percept. Psychophys. **13**, 65–75.
Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.
Miller, C. A., Abbas, P. J., Robinson, B. K., Nourski, K. V., and Rubinstein, J. T. (**2002**). ''The neurophysiological effects of simulated auditory prosthesis stimulation,'' Ninth Quarterly Progress Report, NIH N01-DC-9-2107, Neural Prosthesis Program, NIH, Bethesda, MD.
Moore, B. C. J., and Peters, R. W. (**1997**). ''Detection of increments and decrements in sinusoids as a function of frequency, increment, and decre-

ment duration and pedestal duration,'' J. Acoust. Soc. Am. **102**, 2954–2965.
Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). ''Intensity discrimination as a function of stimulus level with electric stimulation,'' J. Acoust. Soc. Am. **100**, 2393–2414.
Pfingst, B. E., Burnett, P. A., and Sutton, D. (**1983**). ''Intensity discrimination with cochlear implants,'' J. Acoust. Soc. Am. **73**, 1283–1292.
Pfingst, B. E., and Rai, D. T. (**1990**). ''Effects of level on nonspectral frequency difference limens for electrical and acoustic stimuli,'' Hear. Res. **50**, 43–56.
Pfingst, B. E., Zwolan, T. A., and Holloway, L. A. (**1997**). ''Effects of stimulus configuration on psychophysical operating levels and on speech recognition with cochlear implants,'' Hear. Res. **112**, 247–260.
Scharf, B., Canévet, G., and Ward, L. M. (**1992**). ''On the relationship between intensity discrimination and adaptation,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).
Schwid, H. A., and Geisler, C. D. (**1982**). ''Multiple reservoir model of neurotransmitter release by a cochlear inner hair cell,'' J. Acoust. Soc. Am. **72**, 1435–1440.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
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 Nucleaus cochlear implant,'' J. Acoust. Soc. Am. **87**, 905–907.
Shannon, R. V., and Otto, S. R. (**1990**). ''Psychophysical measures from electrical stimulation of the human cochlear nucleus,'' Hear. Res. **47**, 159–168.
Smith, R. L. (**1977**). ''Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects,'' J. Neurophysiol. **40**, 1098–1112.
Smith, R. L., and Brachman, M. L. (**1982**). ''Adaptation in auditory-nerve fibers: A revised model,'' Biol. Cybern. **44**, 107–120.
Smith, R. L., and Zwislocki, J. J. (**1975**). ''Short-term adaptation and incremental response of single auditory nerve fibers,'' Biol. Cybern. **17**, 169–182.
Tanner, Jr., W. P. (**1956**). ''Theory of recognition,'' J. Acoust. Soc. Am. **28**, 882–888.
Viemeister, N. F. (**1988**). ''Psychophysical aspects of auditory intensity coding,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 213–241.
Viemeister, N. F., and Bacon, S. P. (**1988**). ''Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones,'' J. Acoust. Soc. Am. **84**, 172–178.
Westerman, L. A., and Smith, R. L. (**1988**). ''A diffusion model of the transient response of the cochlear inner hair cell synapse,'' J. Acoust. Soc. Am. **83**, 2266–2276.
Zwicker, E., and Fastl, H. (**1972**). ''On the development of the critical band,'' J. Acoust. Soc. Am. **52**, 699–702.

Perceived naturalness of spectrally distorted speech and music

Brian C. J. Moore[a] and Chin-Tuan Tan
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 30 January 2002; revised 4 April 2003; accepted 7 April 2003)

We determined how the perceived naturalness of music and speech (male and female talkers) signals was affected by various forms of linear filtering, some of which were intended to mimic the spectral ''distortions'' introduced by transducers such as microphones, loudspeakers, and earphones. The filters introduced spectral tilts and ripples of various types, variations in upper and lower cutoff frequency, and combinations of these. All of the differently filtered signals (168 conditions) were intermixed in random order within one block of trials. Levels were adjusted to give approximately equal loudness in all conditions. Listeners were required to judge the perceptual quality (naturalness) of the filtered signals on a scale from 1 to 10. For spectral ripples, perceived quality decreased with increasing ripple density up to 0.2 ripple/ERB_N and with increasing ripple depth. Spectral tilts also degraded quality, and the effects were similar for positive and negative tilts. Ripples and/or tilts degraded quality more when they extended over a wide frequency range (87–6981 Hz) than when they extended over subranges. Low- and mid-frequency ranges were roughly equally important for music, but the mid-range was most important for speech. For music, the highest quality was obtained for the broadband signal (55–16 854 Hz). Increasing the lower cutoff frequency from 55 Hz resulted in a clear degradation of quality. There was also a distinct degradation as the upper cutoff frequency was decreased from 16 845 Hz. For speech, there was a marked degradation when the lower cutoff frequency was increased from 123 to 208 Hz and when the upper cutoff frequency was decreased from 10 869 Hz. Typical telephone bandwidth (313 to 3547 Hz) gave very poor quality.
[DOI: 10.1121/1.1577552]

PACS numbers: 43.66.Lj, 43.38.Md, 43.58.Ry [NFV]

I. INTRODUCTION

It is widely accepted that the perceived quality of signals reproduced by transducers such as loudspeakers or headphones can be strongly affected by the frequency response of the transducers (Toole, 1986a, b; Toole and Olive, 1988). It is also generally accepted that ''smooth'' wideband frequency responses are desirable (Bucklein, 1962; Toole, 1986a, b; Toole and Olive, 1988; Gabrielsson *et al.*, 1990, 1991). However, there have been few studies that have aimed to quantify the relationship between perceived quality and frequency response irregularity and/or bandwidth. This study was designed to provide such information.

Reviews of studies on the perceptual effects of irregularities in frequency response are provided in Toole (1986a, b) and Toole and Olive (1988) and we give only a brief overview here. There have been two general approaches to this issue. Some researchers have used filters to introduce spectral distortions with well-controlled characteristics. Bucklein (1962) investigated the audibility (detectability) of single peaks and dips in frequency response, as a function of the sharpness of the peaks and dips, quantified in terms of their ''Q'' value. He found, using speech and music signals, that broad (low Q) peaks and dips were more easily detected than narrow peaks and dips. Fryer (1977) and Toole and Olive (1988) also studied the effects of single peaks (resonances). Like Bucklein, they found that low-Q resonances were more easily heard than high-Q resonances. They also found that resonances were detected more easily using pink or white noise as the test signal than using speech or music. Toole and Olive (1988) reported that, for impulsive or transient sounds, the addition of reverberation could increase the audibility of medium- and low-Q resonances. Gabrielsson *et al.* (1990) investigated the effect on perceived sound quality of filtering signals in three ways, so as to boost low, medium, or high frequencies by 15–20 dB. A reference ''flat'' response was also used. Speech, music, and noise signals were reproduced via an earphone, and listeners were asked to rate the quality of the reproduced sound in terms of loudness, clarity, spaciousness, brightness, softness/gentleness, nearness, and fidelity. When the reproduction level was relatively high, all of the conditions with boosts in response received lower fidelity ratings than the ''flat'' response condition. However, other results suggest that a mild emphasis of medium to high frequencies can actually lead to increased fidelity ratings (Gabrielsson *et al.*, 1988). Bech (2002) studied the effect of low-frequency cutoff and slope and of ripples in frequency response in the low-frequency range (up to 120 Hz). Subjects judged the magnitude of upper bass and lower bass relative to a fixed reference system, using various pieces of pop music as the signal. The cutoff frequency had a significant effect on the ratings of lower bass and the amount of ripple affected ratings of both lower and upper bass. Studies of this type have not attempted to quan-

[a]Electronic mail: bcjm@cus.cam.ac.uk

tify the relationship between frequency response (smoothness or bandwidth) and perceived fidelity.

Other researchers have relied on variations in frequency response from one loudspeaker to another to explore how frequency response affects perceived sound quality. Gabrielsson *et al.* (1991) obtained subjective ratings of musical sounds reproduced by 18 different high-fidelity loudspeakers. They advanced several hypotheses about the relationship between specific perceptual scales and frequency response. For example, ''clarity'' was hypothesized to be ''favored by a broad frequency range and a smooth response with a certain emphasis on midhigh frequencies,'' while ''fullness'' was hypothesized to be ''favored by a broad frequency range but with more emphasis on lower frequencies.'' They found that their hypotheses were supported better when the loudspeaker frequency responses were measured in the listening room than when they were measured in free field or in a reverberation room. However, no quantitative relationships were established between the frequency-response shapes and the perceptual ratings.

Toole (1986a, b) used natural variations between loudspeakers to examine the relationship between frequency responses and listener preferences. He reported that ''it is possible to see a progressive increase in the smoothness as a function of fidelity ratings. The loudspeakers with lower ratings tend to exhibit more fine structure in the curves which, in the spatially averaged data, indicates the presence of resonances within the loudspeaker system.'' He also emphasized the importance of the directivity pattern of the loudspeakers, and suggested that a good loudspeaker should have as uniform a directivity as possible over a wide frequency range. Again, however, no attempt was made to quantify the relationship between the degree of smoothness of the frequency response and the preferences of the listeners.

There are several limitations of previous studies that have used natural variations in loudspeakers to explore the perceptual consequences of nonflat frequency responses. First, the studies have mostly used high-fidelity loudspeakers, so the deviations from flat responses have been relatively small, typically less than ± 4 dB. Thus, the results are not applicable to many of the transducers that are in common use, such as those in telephones, low-cost earphones, portable radios, public address systems, and computers. Second, the frequency response of loudspeakers can be measured in many different ways (for example, on-axis in an anechoic room, averaged for many directions in an anechoic room, in a reverberation chamber, in typical listening rooms) and there is no clear consensus as to the ''best'' method. Third, loudspeakers can differ in many factors other than their frequency response (e.g., in nonlinear distortion, phase response, and directivity) and it is difficult to isolate the effects of these other factors from the effects of differences in frequency response. Finally, these studies have not provided quantitative estimates of the effect on perceived sound quality of different types of spectral distortions.

The ability to detect changes in the spectral shape of stimuli has been widely studied in the context of profile analysis (Green, 1988). However, studies of this topic have generally required subjects to distinguish between two stimuli differing in spectral shape (and level), and, with a few exceptions (Gockel and Colonius, 1997), the spectral characteristics of the stimuli have been fixed throughout a ''run.'' This is not representative of the situation of listening to speech or music through a transducer with a nonflat frequency response, where the signal itself has a spectrum that changes over time. Also, studies of profile analysis have mainly involved the measurement of thresholds for detection of a change in spectral shape, rather than quantification of the perceptual effect of a change in spectral shape.

The goal of the present study was to explore how the perceived naturalness of music and speech signals was affected by various forms of linear filtering. We wanted to isolate the effects of irregularities in amplitude response from other forms of distortion that can occur in transducers, such as phase distortion and nonlinear distortion. Of course, the overall quality of a transducer is determined by the combined effects of all such factors. Our ultimate goal is to produce a perceptually based model that can predict the effects of all forms of distortion. However, we believe that this can only be achieved by studying the effect of each form of distortion separately. The data presented here are intended to represent a first step towards characterizing the perceptual effect of nonflat amplitude responses.

Some of the types of filtering that we introduced were intended to mimic the spectral distortions that can occur in transducers. These included spectral tilts and band-limiting. We also included spectral ripples that were sinusoidal on an ERB_N scale (see below for details). Although ripples in the frequency response of conventional loudspeakers are not usually regular, some transducers, such as the distributed mode (flat panel) loudspeaker, do show fairly regular ripples in their response (Gontcharov *et al.*, 1999). The artificial regular ripples used here were intended to allow us to characterize the effects of ripple density and center frequency. Some of the filter characteristics were deliberately designed to give rather extreme degradations of perceived naturalness, so as to provide listeners with clear examples of unnatural sound qualities. The sounds were reproduced by a high-quality earphone with a good approximation to a diffuse field characteristic, which was especially selected for its low harmonic and intermodulation distortion. This allowed us to be confident that the main factor affecting the subjective responses was the spectral filtering. However, because no reverberation was present in the filtered stimuli, the results should not be considered as applicable to sounds reproduced via loudspeakers in a normal listening room; rather, the results are applicable to sound reproduction via headphones and earpieces, such as those used in telephones. An important feature of our approach was that all filtering conditions were randomly intermixed. This helped to ensure that the response criteria of the subjects were applied consistently to all the various forms of spectral distortion.

II. METHOD

A. Filtering conditions

For many applications, such as in the forthcoming standard for high-quality telephony, the frequency range of inter-

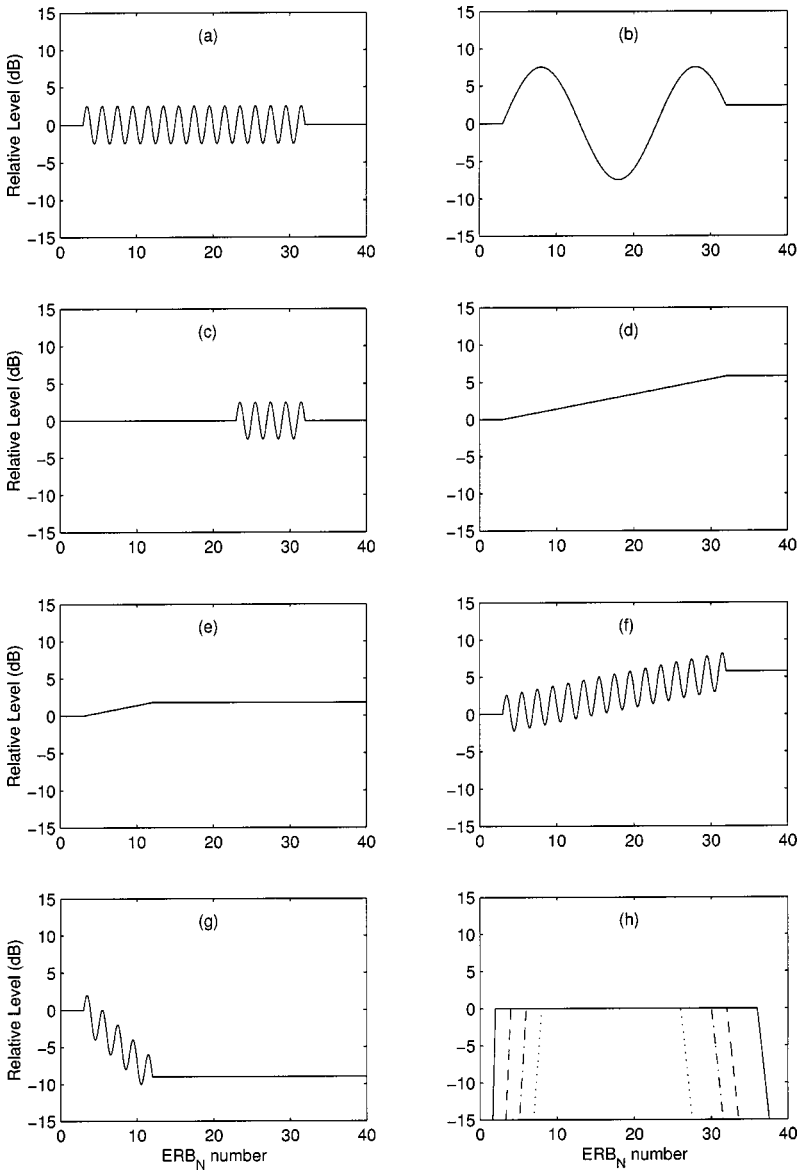


FIG. 1. Schematic illustration of some of the filters used to introduce spectral distortions. The abscissa shows frequency expressed as ERB_N number. (a) Spectral ripple with 5-dB depth and 0.5 ripples/ERB_N, over the range 3–32 ERB_N. (b) Spectral ripple with 1.5-dB depth and 0.05 ripples/ERB_N, over the range 3–32 ERB_N. (c) Spectral ripple with 5-dB depth and 0.5 ripples/ERB_N, over the range 23–32 ERB_N. (d) Positive spectral tilt of 0.2 dB/ERB_N, over the range 3–32 ERB_N. (e) Positive spectral tilt of 0.2 dB/ERB_N, over the range 3–12 ERB_N. (f) Spectral ripple with 5-dB depth and 0.5 ripples/ERB_N, combined with positive spectral tilt of 0.2 dB/ERB_N, both over the range 3–32 ERB_N. (g) Spectral ripple with 5-dB depth and 0.5 ripples/ERB_N, combined with negative spectral tilt of -1 dB/ERB_N, both over the range 3–12 ERB_N. (h) Band-limiting with lower cutoff frequency of 2, 4, 6, or 8 ERB_N and upper cutoff frequency of 26, 30, 32 or 36 ERB_N.

est extends from about 100 to 7000 Hz (frequencies outside that range are not transmitted). As we wished our results to be applicable to telephony, for the set of conditions (1)–(3) below, the spectral distortions were restricted approximately to this frequency range, although the stimuli had a wide bandwidth. We included other conditions with bandpass filtering, including some that simulate current telephone standards (typically 300 to 3400 Hz) and some that present broadband stimuli. In what follows, the characteristics of the filters are described with frequency transformed to ERB_N number, where ERB_N stands for the equivalent rectangular bandwidth of the auditory filter for normal-hearing listeners, when measured at a moderate sound level. The ERB_N scale is a perceptually relevant scale (Moore and Glasberg, 1983; Glasberg and Moore, 1990; Moore, 2003), expressed by the following equation:

$$\text{ERB}_\text{N}\ \text{number} = 21.4 \log_{10}(4.37F + 1), \tag{1}$$

where F is frequency in kHz. This scale is conceptually similar to the Bark scale proposed by Zwicker and co-workers (Zwicker and Terhardt, 1980), although it differs somewhat in numerical values.

Response-modifying filters with the following shapes were used:

(1) Spectral ripples that were sinusoidal in dB on an ERB_N scale. Ripple rates of 0.05, 0.1, 0.2, and 0.5 ripples/ERB_N were used. Ripple depths (peak-to-valley ratios in dB) were 5, 10, and 15 dB. The ripples always started (at the low-frequency end of the range) at a ''zero-crossing.'' The ripples either extended over the range 3–32 ERB_N (87 to 6981 Hz), or were limited to a restricted frequency range, with a flat response outside that range. An example of a wide-range ripple with rate 0.5 ripple/ERB_N and depth 5 dB is shown in Fig. 1(a). In this case, the ripple ended (at the high-frequency end) at a zero crossing. An example of a wide-range ripple with rate 0.05 ripple/ERB_N and depth 15 dB is shown in Fig. 1(b). In this case the ripple ended slightly away from a zero crossing, as only 1.45 ripples were present. The filter responses for frequencies above the ripple range

remained at the value set by the end of the ripple (possible consequences of this are discussed later). Three subranges were used: 3–12, 13–22, and 23–32 ERB_N (corresponding to frequency ranges of 87–606, 701–2224, and 2503–6981 Hz). An example of a ripple with rate 0.5 ripples per ERB_N and depth 5 dB, restricted to the range 23–32 ERB_N, is shown in Fig. 1(c). The use of these subranges allowed us to determine if the perceptual effect of spectral ripples is relatively constant across frequency or varies with frequency. The total number of conditions with ripple was 4 (ripple rates)×3 (ripple depths)×4 (ranges)=48.

(2) Spectral tilts that were linear in dB/ERB_N. Tilts of ±0.1, 0.2, 0.5, and 1 dB/ERB_N were used. The tilts either extended over the frequency range 87 to 6981 Hz, or were limited to a restricted frequency range, with a flat response outside that range. An example of a wide-range positive tilt of 0.2 dB/ERB_N is shown in Fig. 1(d). Three subranges were used: 3–12, 13–22, and 23–32 ERB_N. An example of a positive tilt of 0.2 dB/ERB_N in the subrange 3–12 ERB_N is shown in Fig. 1(e). It should be noted that, for a given tilt in dB/ERB_N, the difference in level between the two ends of the tilt region was three times as great for the wide-range tilt as for any subrange tilt. The total number of conditions with tilt was 8 (tilts)×4 (ranges)=32.

(3) Spectral ripples [as described in (1)] combined with spectral tilts [as described in (2)]. Ripples with rates of 0.1 and 0.5 dB/ERB_N and depths of 5 and 10 dB were combined with tilts of ±0.2 and ±1 dB/ERB_N. Again, the ripples and tilts either extended over the frequency range 87 to 6981 Hz, or were limited to one of the three subranges. An example of full range ripple and a positive tilt of 0.2 dB/ERB_N is given in Fig. 1(f). An example of a ripple combined with a negative tilt of −1 dB/ERB_N, both restricted to the subrange 3–12 ERB_N, is shown in Fig. 1(g). The total number of conditions with both tilt and ripple was 2 (ripple rates)×2 (ripple depths)×4 (tilts)×4 (ranges)=64.

(4) Bandpass filtering with all possible combinations of the following lower and upper cutoff frequencies: lower, 2, 4, 6, and 8 ERB_N (55, 123, 208, and 313 Hz); upper 26, 28, 30, 32, 36, and 40 ERB_N (3547, 4455, 5583, 6981, 10 869, and 16 854 Hz). The response was flat within the passband. Examples of the filter shapes for this condition are given in Fig. 1(h). This gave 24 conditions in total.

B. Filter implementation and stimulus presentation

Each of the 168 filters was implemented digitally, using the overlap-add method (Allen, 1977), which does not introduce any phase distortion. Digital representations of the input signals were obtained directly from CD, using the standard sampling rate of 44 100 Hz. Frames containing 4096 samples (duration =92.9 ms) were Hanning windowed and overlapped by 75%. Thus, frames were updated every 23.2 ms. Two sets of stimuli were used, speech and music, which were evaluated in separate testing sessions. The speech was a concatenation of two sentences, one from a male and one from a female talker, taken from tracks 49 and 50 of the CD ''Sound Quality Assessment Material'' (SQAM) produced by the European Broadcasting Union (www.ebu.ch: materials also available from http://sound.media.mit.edu/mpeg4/audio/sqam/). The same two sentences were used throughout. The overall duration of the two sentences, including the brief pause between them, was 3.1 s. The music was a fragment of jazz (piano, bass and drums) with a relatively constant overall level, taken from a commercial CD (digital recording). The same fragment was used throughout. Its duration was 7.3 s. The speech and music were filtered off-line and the filtered stimuli were stored on computer disk.

The upper panels in Fig. 2 show the power spectral density for the speech (left) and music (right), averaged over the whole duration of the samples. The spectral density is greatest at low frequencies. The lower panels show excitation patterns calculated from the spectra using the method described by Moore *et al.* (1997), and plotted as a function of ERB_N number; the equivalent frequency in Hz is plotted at the top of each panel. This method takes into account the free-field to eardrum transformation (Shaw, 1974), which boosts frequencies around 3 kHz, and the effect of the middle ear, which results in a reduction in the relative level of very low and very high frequencies (Puria *et al.*, 1997; Aibara *et al.*, 2001). The excitation patterns are somewhat flatter than the power spectra, and for both speech and music the peak excitation occurs at an ERB_N number of about 10 (equivalent to about 440 Hz).

The stimuli were replayed to the listener using a 24-bit Lynx 1 sound card, mounted in a PC. The output of the sound card drove Sennheiser HD580 earphones. The same signal was fed to each earpiece. These earphones have a diffuse field response, i.e., they produce at the eardrum of the listener a similar frequency as would be obtained listening in a diffuse sound field. Thus their response at the eardrum shows an increase in the frequency range around 3000 Hz, which reflects the resonance normally produced by the concha and meatus. The earphones were calibrated using a KEMAR manikin (Burkhard and Sachs, 1975), averaging the results for the ''large'' and ''small'' ears. The output of the ear stimulator was connected to a Hewlett-Packard 35670A dynamic signal analyzer. The frequency response was compared to mean of the diffuse field responses of the human ear measured by Shaw (1974), Kuhn (1979), and Killion *et al.* (1987). The response of the earphone was found to match the mean diffuse field response within ±3.5 dB from 30 to 6000 Hz. Above 6000 Hz, the response showed some irregularities which varied depending on which ear was used in KEMAR and also depending on the exact positioning of the earphones on the manikin. Additional measurements using a probe microphone (Etymotic Research ER7C) close to the eardrum of several human individuals showed that the response above 6000 Hz varied from one individual to another, but that the response averaged across individuals was close to the diffuse field response. Such individual variations also occur in the diffuse field responses of humans (Shaw, 1974; Kuhn, 1979; Killion *et al.*, 1987).

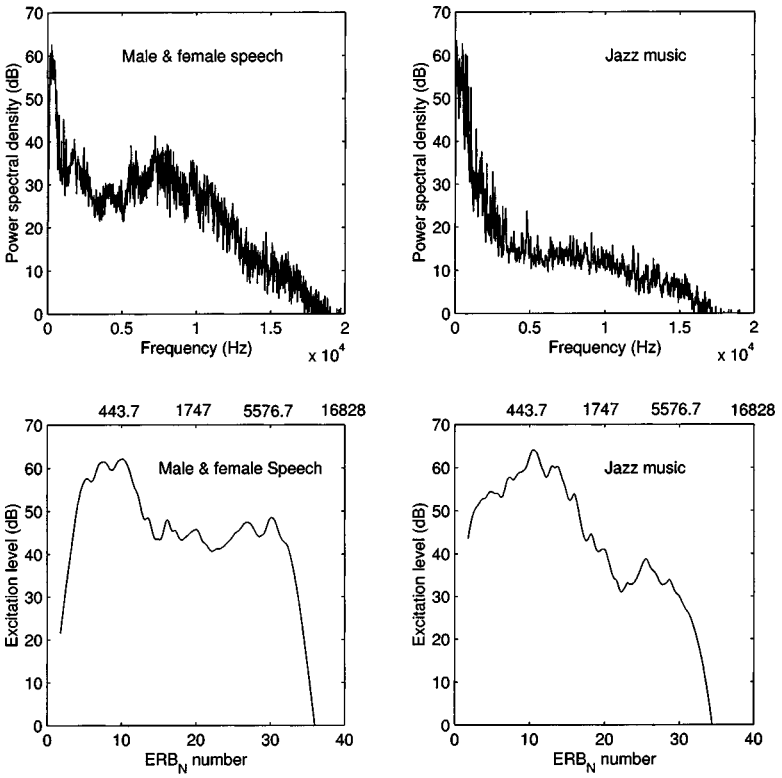


FIG. 2. The upper panels show the mean power spectral density as a function of linear frequency for the speech (left) and music (right). The lower panels show corresponding excitation patterns plotted as a function of ERB_N number.

We also measured the harmonic and intermodulation distortion produced by the earphone. For sinusoidal inputs with frequencies from 100 to 6000 Hz, and for sound levels at the output up to 90 dB SPL, the level of the distortion component corresponding to the second harmonic was always at least 70 dB below the level of the primary component, while the level of the distortion component corresponding to the third harmonic was always at least 84 dB down. Other distortion components were not measurable. For various pairs of primary frequencies, f_1 and f_2, in the range 100 to 7000 Hz, and for levels up to 90 dB SPL, the level of the distortion component at f_2-f_1 was always unmeasurable. For the same conditions, the level of the distortion component at $2f_1-f_2$ was always at least 73 dB lower in level than the primary tones, and was usually unmeasurable. No other distortion components were measurable. We conclude that the nonlinear distortion produced by the earphones was probably below the audible limit (Jacobs and Wittman, 1964; Letowski, 1975; Gabrielsson *et al.*, 1976).

The overall level of each filtered signal was adjusted digitally prior to digital–analog conversion so as to give roughly a constant loudness of 86.4 phons (binaural listening). The required adjustment was calculated using the loudness model of Moore *et al.* (1997). The calculations took into account the diffuse field response of the earphones.

C. Experimental method

In a given session, a listener was tested using either speech or music stimuli. Filtered stimuli were presented in a randomized order. After each stimulus presentation, there was a pause, during which the listener was required to rate the perceived quality on a 10-point scale where ''10'' indicates ''very natural—uncolored'' and ''1'' represents ''very unnatural—highly colored.'' The response categories were displayed on the computer screen, and subjects responded using the mouse to ''click'' on their category of choice. The computer waited indefinitely until a response was made. The next stimulus was presented approximately 1 s after a response was made.

To illustrate the meaning of the descriptors for the categories, before the experiment proper started, samples were presented of wideband signals (55 to 16 854 Hz) without any spectral ripple or tilt; these were described as examples of category 10. Similarly, samples were presented with large amounts of ripple and spectral tilt and described as examples of category 1. In the experiment proper, each subject was tested in four sessions on different days, twice using speech and twice using music. The repeated measurement for each type of stimulus allowed us to assess how consistent the responses of each subject were. Half the subjects were tested first using speech and half using music. In a given session the 168 different conditions were intermixed and presented in random order. Subjects were allowed to pause and rest at any time during a session. A session typically lasted about 1 h for the speech stimuli and 1 h and 10 min for the music stimuli.

D. Subjects

Ten subjects were tested. None had a history of hearing disorders and all had audiometric thresholds better than or equal to 20 dB HL in both ears at all audiometric frequencies from 250 to 8000 Hz. The mean audiometric threshold across subjects and ears was better than 7.3 dB for all audiometric frequencies from 250 to 8000 Hz. Their ages ranged from 15 to 31 years (mean 24, standard deviation 6.2). Subjects were paid for their participation.

TABLE I. Correlation of the mean ratings across sessions for each individual subject with the mean ratings across subjects. Correlations are given separately for each subject for the music stimuli and the speech stimuli. A high correlation for a given subject indicates that the pattern of results for that subject is similar to the pattern found in the mean results.

Subject	1	2	3	4	5	6	7	8	9	10
Music	0.94	0.90	0.90	0.88	0.90	0.89	0.78	0.89	0.85	0.82
Speech	0.92	0.93	0.91	0.84	0.93	0.94	0.93	0.94	0.92	0.87

III. RESULTS

The results for each subject generally showed a very similar pattern across the two test sessions for a given type of signal (speech or music). The overall consistency across test sessions was assessed by calculating the mean score across subjects for each condition and stimulus type, separately for each session, and then calculating the correlation of the scores for the 168 conditions across sessions. The correlations obtained in this way were 0.97 for the music stimuli and 0.97 for the speech stimuli. The very high correlations indicate a high degree of consistency across test sessions.

The pattern of results was also very consistent across subjects. To assess the degree of consistency across subjects, we calculated the mean score for each subject and each condition across the two sessions. We also calculated the mean score across subjects for each condition and stimulus type, including the data for both sessions. Then, for each subject in turn, we calculated the correlation between the scores for that individual subject and the mean scores, over the 168 conditions. The higher the correlation, the more closely the pattern of scores for a given subject resembles that for the group as a whole. The resulting correlations are shown in Table I, separately for music and for speech stimuli. The correlations are generally high, all but one being above 0.82. The high correlations indicate a high degree of consistency across subjects. Hence, in what follows, we focus on the mean results. For simplicity of presentation, most of the figures show the effect of only one type of spectral manipulation.

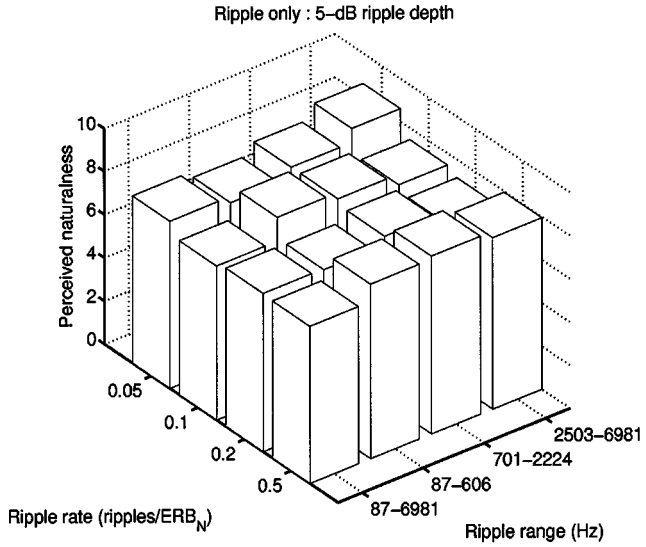
A. Spectral ripples

Figure 3 shows the mean ratings for music (top) and speech (bottom) when processed through filters giving spectral ripples with a depth of 5 dB (peak-valley ratio). Ripples of this magnitude occur quite commonly in the frequency response of good-quality transducers. The *x* axis shows the ripple rate (ripples/ERB_N), the *y* axis shows the ripple range, and the *z* axis shows the mean rating.

Spectral ripples with a depth of 5 dB have only a moderate effect on perceived naturalness. The ripples degrade naturalness more when they extend over a wide frequency range (87–6981 Hz) than when they extend over subranges. Perceived naturalness tends to decrease with increasing ripple rate up to 0.2 ripples/ERB_N.

Figure 4 is similar to Fig. 3, but shows results for a ripple depth of 10 dB. This is the kind of spectral ripple that might be found in the response of lower-quality transducers, such as those in medium quality headphones. The degradation in naturalness produced by this greater ripple depth is clearly evident. Again, the ripples degrade naturalness more when they extend over a wide frequency range (87–6981 Hz) than when they extend over subranges, and perceived naturalness tends to decrease with increasing ripple rate up to 0.2 ripples/ERB_N.

Figure 5 is similar to Fig. 3, but shows results for a ripple depth of 15 dB. This is the kind of spectral ripple that might be found in the response of poor quality earpieces and headphones. This large ripple depth produces substantial degradations of naturalness, and for the wide-range ripple the



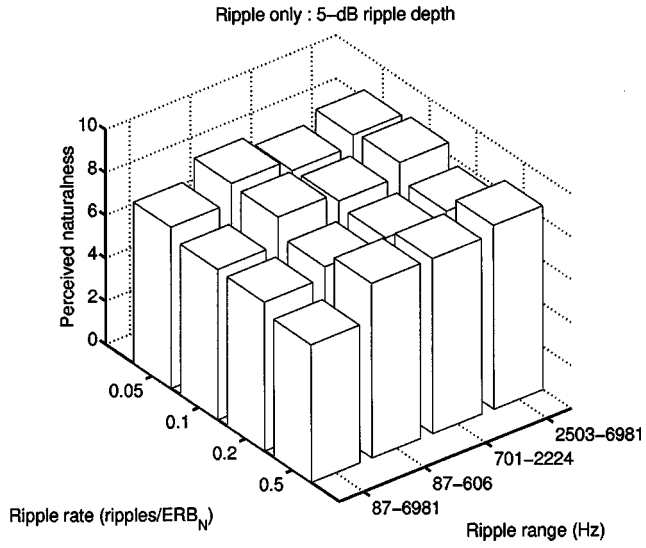
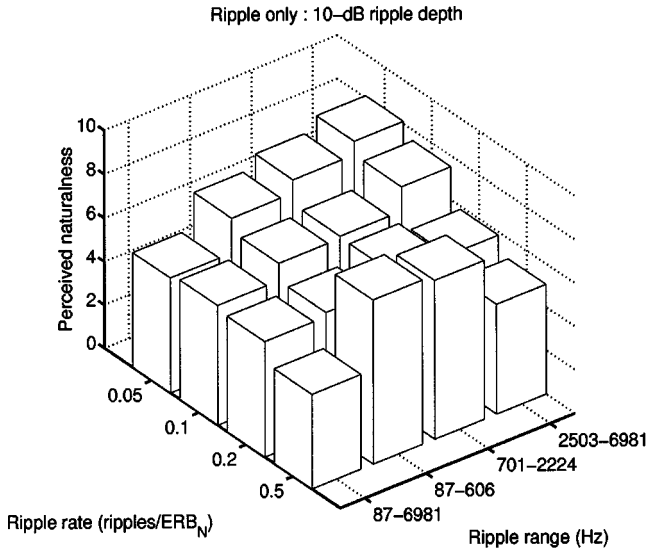
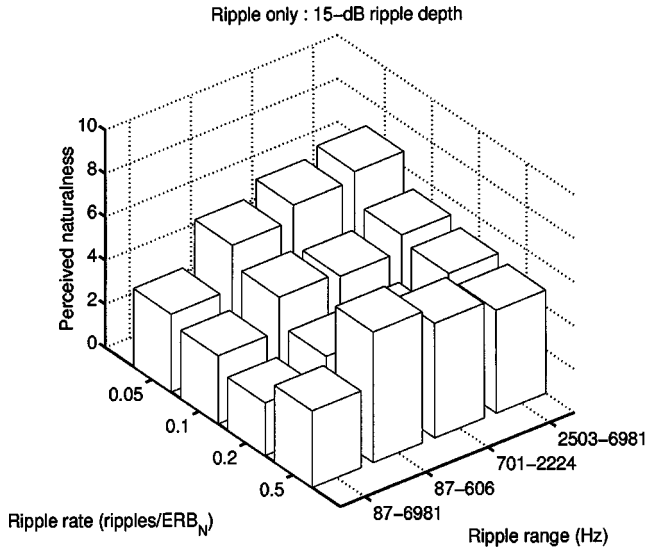


FIG. 3. Mean ratings for music (top) and speech (bottom) when processed through filters giving spectral ripples with a depth of 5 dB. The *x* axis shows the ripple rate (ripples/ERB_N), the *y* axis shows the ripple range, and the *z* axis shows the mean rating.





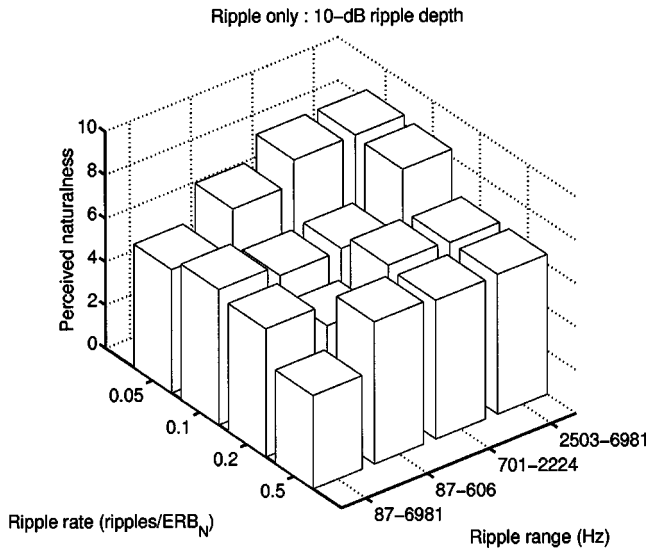


FIG. 4. As Fig. 3, but for a ripple depth of 10 dB.

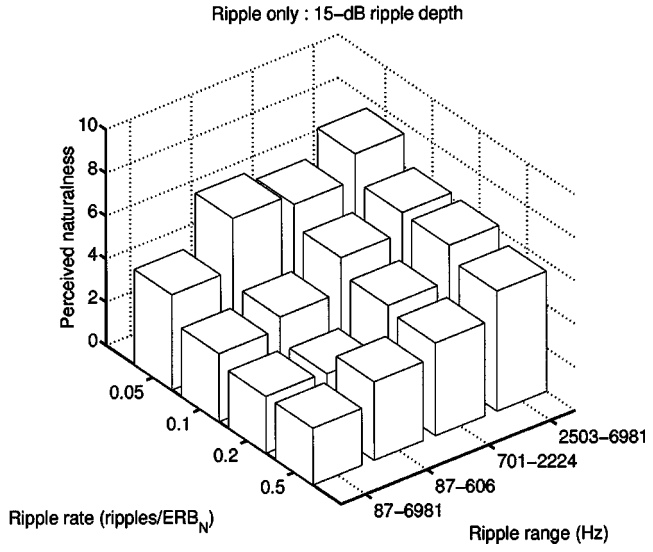


FIG. 5. As Fig. 3, but for a ripple depth of 15 dB.

ratings decrease to values between 2 and 3 for the higher ripple densities. For the speech signal, ripple over the highest frequency subrange produces less degradation than ripple over the low- and mid-frequency subranges.

To assess the statistical significance of the observed effects, the data were subjected to within-subject analyses of variance (ANOVAs). The data from the two test sessions for a given type of stimulus (speech or music) were treated as replications. The least-significant differences test (Snedecor and Cochran, 1967) was used to perform *posthoc* comparisons of means for specific conditions. We performed separate ANOVAs for the music and speech signals, each with factors ripple depth (three values), ripple rate (four values), and ripple range (four values).

For both music and speech, all main effects were significant at $p<0.001$. For music, the mean ratings were 7.7, 6.2, and 4.6 for ripple depths of 5, 10, and 15 dB, respectively. For speech, the corresponding ratings were 7.8, 6.4, and 4.7. For both speech and music, all pairwise comparisons were significant at $p<0.001$. For music, the mean ratings were 6.8, 6.1, 5.5, and 6.2 for ripple rates of 0.05, 0.1, 0.2, and 0.5 ripples/ERB$_N$, respectively. For speech, the corresponding rates were 7.2, 6.3, 5.7, and 5.9. In both cases, the ratings did not differ significantly for rates of 0.1 and 0.5, but all other pairwise comparisons were significant at $p<0.01$. Thus, naturalness at first decreased with increasing ripple rate, but then increased again.

For music, the wide-range ripples gave significantly lower naturalness ratings than all of the subrange ripples ($p<0.001$), but there was no significant effect of subrange center frequency. For speech, wide-range ripples had the greatest deleterious effect, and ripple in subranges produced an effect which decreased with increasing center frequency; the mean rating was 5.2 for the wide-range ripple (3–32 ERB$_N$) and 6.1, 6.6, and 7.3 for the subranges 3–12, 13–22, and 23–32 ERB$_N$, respectively. All pairwise comparisons were significant at $p<0.002$. All of the two-way interactions and the three-way interaction were significant at $p<0.001$.

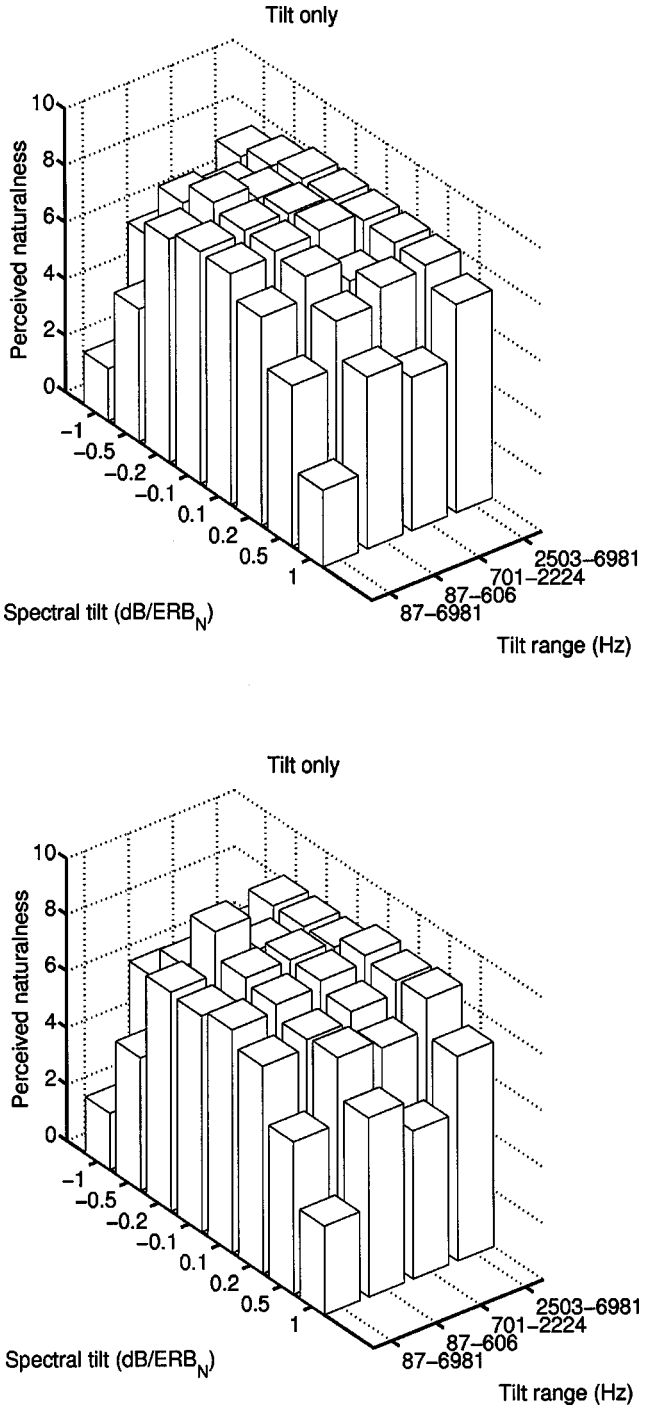


FIG. 6. Mean ratings for music (top) and speech (bottom) with spectral tilts differing in amount and direction and applied over different ranges. The *x* axis shows the spectral tilt (dB/ERB$_N$), the *y* axis shows the tilt range, and the *z* axis shows the mean rating.

However, no interaction accounted for more than 2.5% of the variance in the data, so we will not discuss the interactions further.

B. Spectral tilt

Figure 6 shows results for music (top) and speech (bottom) signals with spectral tilts differing in amount and direction and applied over different ranges. The *x* axis shows the spectral tilt (dB/ERB$_N$), the *y* axis shows the tilt range, and the *z* axis shows the mean rating. Spectral tilts (either upwards or downwards) degrade naturalness, especially when they are applied over the whole frequency range. For the wise-range tilts of large magnitude, the ratings reach values around 2, for both positive and negative tilts. However, tilts restricted to the high-frequency range (2503–6981 Hz) have relatively little perceptual effect, until they are rather extreme.

Separate within-subjects ANOVAs were conducted for the music and speech stimuli, with factors direction of tilt (positive or negative), tilt magnitude, and tilt range. For both music and speech, the main effect of tilt direction was not significant, but the other two main effects were significant at $p<0.001$. For music, the mean rating was 5.8 for the wide-range tilt (3–32 ERB$_N$) and 7.5, 7.2, and 7.8 for the subranges 3–12, 13–22, and 23–32 ERB$_N$, respectively. The mean ratings did not differ significantly for the 3–12 and 13–22 ERB$_N$ subranges. All other pairwise comparisons were significant at $p<0.05$. For speech, the mean rating was 5.7 for the wide-range tilt (3–32 ERB$_N$) and 7.6, 7.1, and 7.9 for the subranges 3–12, 13–22, and 23–32 ERB$_N$, respectively. All pairwise comparisons were significant at $p<0.01$. Thus, mid-range tilt had a slightly perceptual effect than low- or high-range tilt. This greater effect may have occurred partly because, for the mid-range tilt, there were large frequency ranges below and above the tilt region (i.e., below 701 Hz and above 2224 Hz) across which there was a difference in relative level. For the low-range tilt, the ''plateau'' region on the low-frequency side occurred only below 87 Hz, while for the high-range tilt the plateau on the high-frequency side occurred only above 6981 Hz.

C. Combined tilt and ripple

Figure 7 shows the results for speech and music signals with a positive or negative tilt of 1 dB/ERB$_N$ combined with spectral ripples of 10 dB depth. In each panel, the *x* axis shows the ripple rate, the *y* axis shows the ripple range, and the *z* axis shows the mean rating. A large upward or downward tilt combined with a spectral ripple applied over the whole frequency range leads to a severe degradation of naturalness, for both low and high ripple densities; mean ratings approach 2 for these conditions. However, tilt and ripple restricted to the high-frequency range have little perceptual effect. Tilt and ripple in the mid-range tend to have the largest perceptual effect for speech, but for music the low- and mid-ranges are equally important.

Separate within-subjects ANOVAs were conducted for the music and speech stimuli, with factors direction of tilt, tilt magnitude, ripple depth, ripple rate, and range. For both music and speech, the main effect of tilt direction was not significant. Also, the main effect of ripple depth was not significant. Apparently, when a large tilt is present, its effects ''swamp'' those of ripple depth. The effect of ripple rate was not significant, which is consistent with the findings for ripple only for the two rates used here. The main effect of range was significant at $p<0.001$ for both speech and music. For music, the mean ratings were 5.0 for the wide-range tilt and ripple (3–32 ERB$_N$) and 6.9, 6.8, and 7.6 for the subranges 3–12, 13–22, and 23–32 ERB$_N$, respectively. The

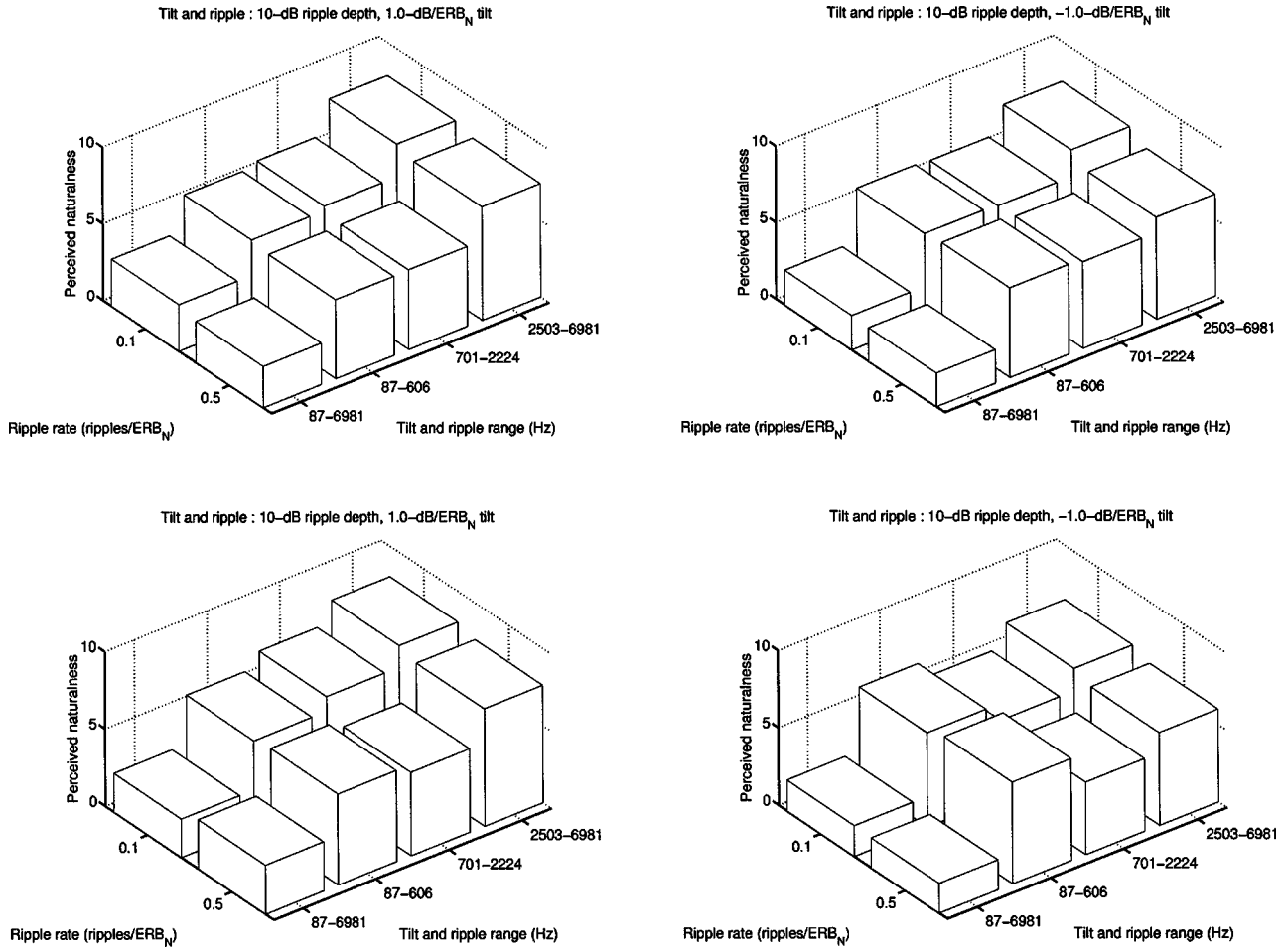


FIG. 7. Mean ratings for music (top) and speech (bottom) with a positive (left) or negative (right) tilt of 1 dB/ERB$_N$ combined with spectral ripples of 10-dB depth. The *x* axis shows the ripple rate, the *y* axis shows the tilt/ripple range, and the *z* axis shows the mean rating.

ratings did not differ significantly for ranges of 3–12 and 13–22 ERB$_N$, but all other pairwise comparisons were significant at $p<0.001$. For speech, the mean ratings were 4.8 for the wide-range tilt and ripple and 7.3, 6.8, and 7.7 for the subranges 3–12, 13–22, and 23–32 ERB$_N$, respectively. All pairwise comparisons were significant at $p<0.01$. For both music and speech, the results show that wide-range tilt and ripple has a greater effect than for any subrange, and that tilt and ripple at high frequencies has a smaller effect than at mid or low frequencies.

D. Bandlimiting

Figure 8 illustrates the effects of bandlimiting the stimuli. The *x* axis shows the high-frequency cutoff, the *y* axis shows the low-frequency cutoff, and the *z* axis shows the mean rating. For music (top), the highest naturalness is obtained for the broadband signal (55–16 854 Hz). Increasing the lower cutoff frequency from 55 to 125 Hz results in a clear degradation of naturalness. There is also a reasonably progressive decrease in naturalness as the upper cutoff frequency is decreased from 16 854 to 3547 Hz. For speech (bottom), there is little effect of increasing the lower cutoff frequency from 55 to 123 Hz, but a marked degradation when the lower cutoff frequency is increased from 123 to 208 Hz. Also there is a progressive decrease in naturalness as the upper cutoff frequency is decreased from 10 869 to 3547 Hz. Typical telephone bandwidth (313 to 3547 Hz) gives very poor naturalness for both speech and music.

Within-subjects ANOVAs were conducted separately for the music and speech signals, with factors lower cutoff frequency and upper cutoff frequency. For both music and speech, both main effects and the two-way interaction were significant at $p<0.001$. When the lower cutoff frequency was low (55 or 123 Hz), changes in the upper cutoff frequency had a large effect, while when the lower cutoff frequency was high (313 Hz), changes in the upper cutoff frequency had only a small effect; in the latter case ratings were consistently very low. Similarly, when the upper cutoff frequency was high (10 869 or 18 854 Hz), changes in the lower cutoff frequency had a large effect, while when the upper cutoff frequency was lower (4455 or 3547 Hz), changes in the lower cutoff frequency had only a small effect; again, ratings in the latter case were consistently very low.

The results for both music and speech indicate that poor naturalness introduced by a high lower cutoff frequency cannot be compensated by changing the upper cutoff frequency; and poor naturalness introduced by a low upper cutoff frequency cannot be compensated by changing the lower cutoff frequency.

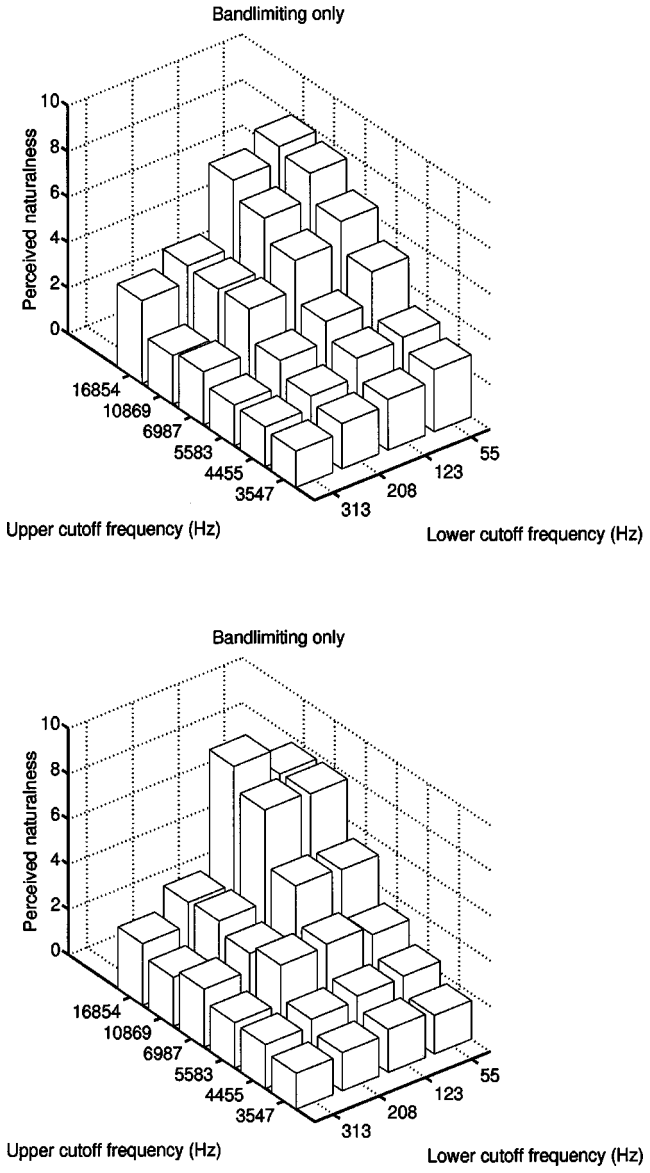


FIG. 8. Mean ratings for band-limited music (top) or speech (bottom) stimuli. The *x* axis shows the high-frequency cutoff, the *y* axis shows the low-frequency cutoff, and the *z* axis shows the mean rating.

IV. DISCUSSION

It is possible that some of the results for the stimuli with spectral ripple were influenced by the fact that the ripple did not always end (on its high-frequency side) at 0 dB; the filter response for frequencies above the ripple range remained at the value set by the end of the ripple. Such an effect was present for all ripple rates except 0.5 ripples/ERB_N. For the 15-dB ripple depth, the "offset" was +2.3, −4.4, and −7.1 dB for the ripple rates of 0.05, 0.1, and 0.2 ripples/ERB_N, respectively. For smaller ripple depths, the offset was correspondingly smaller. The largest offset was somewhat smaller than the −9-dB offset produced by a spectral tilt of −1 dB/ERB_N over a 9-ERB range. While tilts of −1 dB/ERB_N did produce a clear degradation in naturalness (ratings were between 5 and 7, depending on the subrange over which the tilt was applied), the degradation was smaller than produced by 15-dB ripples at a rate of 0.2 ripples/ERB_N, which were in the range 3.5 to 5. Thus, the dominant effect for the ripples was probably due to the ripples *per se*, but the offsets may have played some role.

The effect of the offsets may partly account for the finding that naturalness decreased with increasing ripple density up to 0.2 ripples/ERB_N, but then increased slightly at 0.5 ripples/ERB_N; the ripple rate of 0.5 ripples/ERB_N was the only one for which there was no offset. However, the nonmonotonic effect of ripple rate occurred even when the ripple depth was only 5 dB, when the maximum offset was only −2.4 dB. Thus, some other factor was probably involved. The most likely one is that the limited resolution of the auditory filters has the effect of smoothing fine ripples in the spectrum, so that the representation of ripples in the excitation pattern is reduced.

Our procedure of equating the loudness of all of the filtered stimuli meant that the overall level in different frequency regions differed across stimuli. For example, for stimuli filtered with a positive spectral tilt, the overall level at high frequencies was greater than for the undistorted stimuli, while the level at low frequencies was lower. It is, of course, impossible to introduce differences in spectral shape without introducing differences in overall level in some frequency regions. Subjects may partly have based their judgments on the overall level in specific frequency regions, rather than on the relative spectral shape. However, subjects were specifically asked to judge overall naturalness and not loudness, so we feel that this is unlikely.

Generally, the ratings for the speech stimuli showed a very similar pattern to those for the music stimuli. This is confirmed by the fact that the correlation of the mean ratings for the two stimulus types, taken across all 168 conditions, was 0.96. This high correlation again confirms that the mean ratings are highly reliable and reproducible. We believe that this high reliability is linked to the method that we employed; the use of a random mixture of all conditions, and the inclusion of stimuli with very marked spectral distortion allowed subjects to maintain stable criteria, and helped to give consistent criteria across subjects. It is noteworthy that the mean rating for any condition was never above 8.6 for music and 9.1 for speech, and was never below 1.55 for speech or music. This reflects the well-known reluctance of subjects to use the extremes of the available range of responses when making subjective judgements (Poulton, 1979).

There were some minor differences between the results for the speech and music stimuli. Specifically, for speech, the band-limiting conditions revealed no significant effect of increasing the lower cutoff frequency from 55 to 123 Hz or of decreasing the upper cutoff frequency from 16 854 to 10 869 Hz, while for music the corresponding changes did have significant effects.

There were also some small differences in the relative importance of the different frequency subranges for speech and music. It was always the case, for both speech and music, that wide-range tilt and/or ripple had greater deleterious effects on the naturalness ratings than tilt and/or ripple in any subrange. However, for speech, tilt or tilt plus ripple had a greater effect in the mid-range than in the low range, while

for music the effects were similar for the low- and mid-ranges. For both speech and music, ripple and/or tilt had a smaller effect in the high-range than in the mid- or low-ranges.

It is noteworthy that, for the music stimuli, the mean score with the smallest amounts of tilt (± 0.1 dB/ERB$_N$) was slightly higher at about 8.2 than the mean score in the condition where stimuli were band-limited to the range 55 to 16 854 Hz, but with a flat response in the passband (mean score=7.9). This is probably a consequence of the fact that in the conditions with tilt, the stimuli actually had a wider bandwidth (about 30 to 20 000 Hz) than in the ''broadband'' condition with band-limiting. Probably, frequencies below 55 Hz led to slightly improved naturalness for the music stimuli in the conditions with small amounts of tilt.

Our results provide no support for the claim of Gabrielsson *et al.* (1988) that a mild emphasis of medium to high frequencies can lead to increased fidelity ratings. For both music and speech, scores were almost identical for positive and negative tilts of 0.1 dB/ERB$_N$ over the range 13–22 ERB$_N$ (701–2224 Hz) or 23–32 ERB$_N$ (2503–6981 Hz).

V. CONCLUSIONS

We determined how the perceived naturalness of music and speech signals was affected by various forms of linear filtering, some of which were intended to mimic the spectral ''distortions'' typically introduced by transducers such as microphones, loudspeakers, and earphones. Ratings of naturalness were obtained on a scale from 1 to 10, where 10 represents the most natural and 1 the least. For each stimulus type (music and speech), each of the 168 conditions was rated once in a given test session. Each of ten subjects was tested twice in different sessions for each stimulus type. The main conclusions are as follows:

(1) The ratings were highly consistent across test sessions. The correlation of mean scores across the two sessions was 0.97 for music and 0.97 for speech.
(2) The pattern of scores across conditions (with a few minor exceptions noted below) was very similar for music and speech; the correlation of scores for the two stimulus types was 0.96.
(3) For spectral ripples alone, perceived naturalness decreased progressively as ripple depth was increased. Ratings decreased with increasing ripple rate up to 0.2 ripples/ERB$_N$ and then increased slightly. For both speech and music, the ripples had a greater effect when they occurred over a wide frequency range than when they occurred in subranges. For speech, ripples in the high-frequency range had a smaller effect than ripples in the mid- or low-range.
(4) For tilts alone, ratings decreased progressively with increasing tilt magnitude. The effects were similar for positive and negative tilts. For both speech and music, the tilts had a greater effect when they occurred over a wide frequency range than when they occurred in subranges. This is probably largely a consequence of the fact that, for a fixed tilt value in dB/ERB$_N$, the difference in level between the start and end of the tilt was greater for the wide-range tilt than for the subrange tilts. For both music and speech, tilts in the high-frequency range had a smaller effect than tilts in the mid- or low-range. For speech, tilts in the low-frequency range had a smaller effect than tilts in the mid-frequency range.
(5) For combined tilts and ripples, the degradations in naturalness were greater when the tilts and ripples occurred over a wide frequency range than when they occurred in subranges. For both music and speech, tilts and ripples in the high-frequency range had a smaller effect than tilts in the mid- or low-range.
(6) For band-limited stimuli with a flat response within the passband, there were large effects of manipulating the band edge frequencies.
(7) For music, the highest naturalness was obtained for the broadband signal (55–16 854 Hz). Increasing the lower cutoff frequency from 55 Hz resulted in a clear degradation of naturalness. There was also a distinct degradation as the upper cutoff frequency was decreased from 16 845 Hz. Typical telephone bandwidth (313 to 3547 Hz) gave very poor naturalness.
(8) For speech, there was little effect of increasing the lower cutoff frequency from 55 to 123 Hz, but further increases led to a degradation of naturalness. There was also little effect of decreasing the upper cutoff frequency from 16 854 to 10 869 Hz, but further decreases led to a degradation of naturalness. Typical telephone bandwidth (313 to 3547 Hz) gave very poor naturalness.
(9) For both speech and music, when the lower cutoff frequency was relatively high, changes in the upper cutoff frequency had only a small effect. Similarly, when the upper cutoff frequency was low, changes in the lower cutoff frequency had only a small effect.

ACKNOWLEDGMENTS

This project was supported by Nokia Corporation. We thank Nick Zacharov, Brian Glasberg, and Michael Stone for their help with various aspects of this work. We also thank Michael Stone, Thomas Baer, Neal Viemeister, and three anonymous reviewers for helpful comments on an earlier version of this paper.

Aibara, R., Welsh, J. T., Puria, S., and Goode, R. L. (**2001**). ''Human middle-ear sound transfer function and cochlear input impedance,'' Hear. Res. **152**, 100–109.
Allen, J. B. (**1977**). ''Short term spectral analysis, synthesis and modification by discrete Fourier transform,'' IEEE Trans. Acoust., Speech, Signal Process. **25**, 235–238.
Bech, S. (**2002**). ''Requireents for low-frequency sound reproduction, Part I: The audibility of changes in passband amplitude ripple and lower system cutoff frequency and slope,'' J. Audio Eng. Soc. **50**, 564–580.
Bucklein, R. (**1962**). ''Hörbarkeit von Unregelmässigkeiten in Frequenzgängen bei akustischer Übertragung,'' Frequenz **16**, 103–108.
Burkhard, M. D., and Sachs, R. M. (**1975**). ''Anthropometric manikin for acoustic research,'' J. Acoust. Soc. Am. **58**, 214–222.
Fryer, P. A. (**1977**). ''Loudspeaker distortions: Can we hear them?'' Hi-Fi News Rec. Rev. **22**, 51–56.
Gabrielsson, A., Hagerman, B., Bech-Kristensen, T., and Lundberg, G. (**1990**). ''Perceived sound quality of reproductions with different frequency responses and sound levels,'' J. Acoust. Soc. Am. **88**, 1359–1366.
Gabrielsson, A., Lindström, B., and Till, O. (**1991**). ''Loudspeaker frequency response and perceived sound quality,'' J. Acoust. Soc. Am. **90**, 707–719.

Gabrielsson, A., Nyberg, P. O., Sjögren, H., and Svensson, L. (**1976**). ''Detection of amplitude distortion by normal hearing and hearing impaired subjects,'' Karolinska Institute, Technical Audiology **TA83**, 1–20.
Gabrielsson, A., Schenkman, B. N., and Hagerman, B. (**1988**). ''The effects of different frequency responses on sound quality judgments and speech intelligibility,'' Hear. Res. **31**, 166–177.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Gockel, H., and Colonius, H. (**1997**). ''Auditory profile analysis: Is there perceptual constancy for spectral shape for stimuli roved in frequency?'' J. Acoust. Soc. Am. **102**, 2311–2315.
Gontcharov, V. P., Hill, N., and Taylor, V. (**1999**). ''Measurement Aspects of Distributed Mode Loudspeakers,'' in *AES 106th Convention* (Munich).
Green, D. M. (**1988**). *Profile Analysis* (Oxford U.P., Oxford).
Jacobs, J. E., and Wittman, P. (**1964**). ''Psychoacoustics, the determining factor in stereo disc distortion,'' J. Audio Eng. Soc. **12**, 115–123.
Killion, M. C., Berger, E. H., and Nuss, R. A. (**1987**). ''Diffuse field response of the ear,'' J. Acoust. Soc. Am. Suppl. 1 **81**, S75.
Kuhn, G. (**1979**). ''The pressure transformation from a diffuse field to the external ear and to the body and head surface,'' J. Acoust. Soc. Am. **65**, 991–1000.
Letowski, T. (**1975**). ''Difference limen for nonlinear distortion in sine signals and musical sounds,'' Acustica **34**, 106–110.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).
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., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.
Poulton, E. C. (**1979**). ''Models for the biases in judging sensory magnitude,'' Psychol. Bull. **86**, 777–803.
Puria, S., Rosowski, J. J., and Peake, W. T. (**1997**). ''Sound-pressure measurements in the cochlear vestibule of human-cadaver ears,'' J. Acoust. Soc. Am. **101**, 2754–2770.
Shaw, E. A. G. (**1974**). ''Transformation of sound pressure level from the free field to the eardrum in the horizontal plane,'' J. Acoust. Soc. Am. **56**, 1848–1861.
Snedecor, G. W., and Cochran, W. G. (**1967**). *Statistical Methods* (Iowa U. P., Ames).
Toole, F. E. (**1986a**). ''Loudspeaker measurements and their relationship to listener preferences: Part 1,'' J. Audio Eng. Soc. **34**, 227–235.
Toole, F. E. (**1986b**). ''Loudspeaker measurements and their relationship to listener preferences: Part 2,'' J. Audio Eng. Soc. **34**, 323–348.
Toole, F. E., and Olive, S. E. (**1988**). ''The modification of timbre by resonances: Perception and measurement,'' J. Audio Eng. Soc. **36**, 122–142.
Zwicker, E., and Terhardt, E. (**1980**). ''Analytical expressions for critical band rate and critical bandwidth as a function of frequency,'' J. Acoust. Soc. Am. **68**, 1523–1525.

Precedence-effect thresholds for a population of untrained listeners as a function of stimulus intensity and interclick interval

Kourosh Saberi[a)]
Division of Biology, Caltech, Pasadena, California 91125

Joseph V. Antonio
Department of Psychology, California State University, Los Angeles, California 90032

(Received 12 September 2001; revised 2 April 2003; accepted 4 April 2003)

Data are reported from 127 untrained individuals under lag- and single-click conditions in a precedence-effect task. In experiment I, each subject completed ten runs in a two-interval forced-choice design under a lag-click condition and three runs under a single-click condition. The cue to be discriminated was an interaural time difference (ITD). Stimuli were 125-μs rectangular pulses and the interclick interval (ICI) was 2 ms. Subjects were randomly assigned to three groups of approximately 30. Each group was tested at one stimulus intensity (43, 58, or 73 dB). Mean threshold within each group was greater than 500 μs for lag-click ITD conditions, although substantial intersubject variability and a clear effect of stimulus intensity on lag-click ITD thresholds were observed, with lower thresholds for higher intensities. In experiment II, the ICI was varied from 0.3 to 10 ms, and thresholds were obtained from groups of approximately 20 untrained subjects. Data were also collected from three highly experienced observers as a function of ICI. The best naïve subject produced mean thresholds near, but not as low as those obtained from experienced subjects. Analysis of adaptive-track patterns revealed abrupt irregularities in threshold tracking, consistent with either losing the cue or listening to the wrong cue in an ambiguous stimulus.
 [DOI: 10.1121/1.1578079]

PACS numbers: 43.66.Pn, 43.66.Mk, 43.66.Qp [LRB]

I. INTRODUCTION

The precedence effect refers to a variety of onset-dominance phenomena in localization of auditory stimuli (Wallach *et al.*, 1949; Haas, 1949). One defining feature of this effect is a decrease in the ability to perceive spatial information in a sound preceded by another, usually transient, sound. The precedence effect holds for binaural as well as monaural conditions, including along the vertical and front-back axes (Blauert, 1971, 1997; Rakerd *et al.*, 2000). In spite of an extensive history of theoretical (Lindemann, 1986a, b; Zurek, 1980, 1987; Freyman *et al.*, 1997; Saberi, 1996; Hartung and Trahiotis, 2001), neurophysiological (Cranford and Oberholtzer, 1976; Yin and Litovsky, 1995; Mickey and Middlebrooks, 2001), applied (Blauert, 1989; Muncey *et al.*, 1953), and even clinical (Hochster and Kelly, 1981) research, the mechanisms underlying the precedence effect are not well understood. For reviews see Gardner (1968), Zurek (1987), and Litovsky *et al.* (1999).

Some evidence suggests that the precedence effect is at least partially governed by low-level noncortical processes (Hafter *et al.*, 1983, 1988; Hafter and Dye, 1983; Zurek, 1987; Hartung and Trahiotis, 2001). One such view postulates neural inhibition as a mechanism for reduction of spatial cues (Lindemann, 1986a, b; Hafter *et al.*, 1988), particularly when the lead and lag waveforms contain overlapping spectral energies (Shinn-Cunningham *et al.*, 1995; Saberi, 1996). Another approach suggests that peripheral processes including interactions within auditory filters and the interaural phase spectrum can account for many features of the precedence effect, particularly when transient stimuli are employed (Saberi and Perrott, 1995; Tollin and Henning, 1998, 1999; Hartung and Trahiotis, 2001; Zurek and Saberi, 2003). Other evidence, however, points to a complex process involving high-order influences. One striking example is the Clifton effect (1987); Clifton and her colleagues have demonstrated a failure of the precedence effect when the position of the lead and lag sources are reversed between successive presentations (Clifton and Freyman, 1989). These experiments have shown that when the effect fails, it takes several seconds for it to be built up, implicating a slow centralized process. Further evidence against a hard-wired explanation is provided by studies that show learning (Saberi and Perrott, 1990) and top-down influences when cross-correlation cues are ambiguous (Zurek and Saberi, 2003; Saberi and Perrott, 1995; Freyman *et al.*, 1997). Related phenomena in localization, such as the Franssen effect (1960) which demonstrates onset dominance in reverberant environments lasting tens of seconds, are also thought to be driven by such cognitive factors as plausibility (Hartmann and Rakerd, 1989).

Because these diverse findings are inconsistent with a single explanation, current views consider the precedence effect to incorporate several onset phenomena, which in turn has necessitated multiple approaches to its study (Blauert, 1997; Djelani and Blauert, 2001); The reader is referred to Blauert and Col (1991) for a discussion of irregularities in

[a)]Current address: Department of Cognitive Sciences, University of California, Irvine, CA 92697. Electronic mail: kourosh@uci.edu

defining the precedence effect. The current study uses a population approach to study precedence in which we examine thresholds from a large number of untrained listeners. The study was motivated by observations of large variability across threshold estimates in several experiments, particularly from naive subjects, that appear statistically nonstationary and difficult to attribute to sensory inhibition (Saberi and Perrott, 1990; Zurek and Saberi, 2003; Tollin and Henning, 1998, 1999), and by discrepant findings between Saberi and Perrott (1990) and Litovsky *et al.* (2000) related to the effects of practice on lag-click ITD discrimination.

In the current study, we examined population thresholds for two-transient dichotic stimuli. The population was composed of experimentally naïve young college students. This approach allowed the establishment of a databank as well as a baseline to compare with thresholds from experienced subjects. In the current experiments, thresholds were obtained from 127 individuals, grouped randomly into sets of approximately 20 to 30. We selected this group size because it provided sufficient statistical power for estimation of group parameters (Pitman, 1993; Hays, 1981). For each group, we obtained interaural delay thresholds as a function of interclick interval (ICI) and stimulus intensity because intensity has previously been shown to affect lag-click thresholds (Goverts *et al.*, 2000). Finally, we explored irregularities in patterns of responses observed in adaptive tracks. Analysis of tracks revealed unstable psychometric functions, predominately for naïve listeners, but also for experienced subjects, implicating a possible influence of conflicting stimulus cues.

II. EXPERIMENT I: INTERAURAL DELAY THRESHOLDS FOR SINGLE AND LAG CLICKS AS A FUNCTION OF STIMULUS INTENSITY

A. Method

Stimuli were 125-μs rectangular pulses generated by a Dell PC (OptiPlex GX1) and presented through 16-bit digital-to-analog converters (Sound Blaster Live, Milpitas, CA) at a sampling rate of 40 kHz and were low-pass filtered at 20 kHz. Subjects listened to stimuli over Sony (MDR-V1) headphones in an acoustically isolated steel chamber (Industrial Acoustics Company; interior dimensions of $1.8\times1.9\times2\ m^3$). The level of a single pulse was calibrated to 43, 58, or 73 dB (A weighted, slow time average), depending on the experimental condition, using a 6-cc coupler, 0.5-in. microphone (Brüel & Kjær, Model 4189), and a modular precision sound analyzer (Brüel&Kjær, Type 2260).[1] Measurements of the headphone outputs, using a 6-cc coupler, 0.5-in. microphone (B&K), a conditioning amplifier (Nexus, B&K), and an analog-to-digital (A/D) converter (Sound Blaster), showed that the pulse spectrum was linear for the three intensities tested (43, 58, and 73 dB). The timing between pulses and between channels, as well as the level between channels, were checked for accuracy at each SPL with a dual-channel digital storage oscilloscope (Tektronix, Model TDS210) and the microphone assembly described above. Measurements showed that for all three stimulus intensities, the interclick interval for a two-click train with an ICI of 2 ms, measured at the *output* of the headphones, was unaffected by the chosen intensities and the headphone transfer function. Timing between channels was also checked at the output of the headphones by placing a microphone equidistant from the two headphone channels and recording dichotic pulses with different delays between channels. This measurement showed that between-channel timing was precise at the three tested intensities.

Subjects were untrained young college students (ages 18–22) who were recruited from campus advertisements and were paid an hourly wage for their participation. In addition to an hourly wage, subjects were told that the individual with the lowest overall threshold within their respective group (of approximately 30 subjects) would receive a financial bonus. All subjects had normal hearing based on self-report. All subjects were asked if they had a head cold or congestion, and, if so, were rescheduled. Subjects listened to each task in an initial pilot run in the presence of the experimenter for several trials until the experimenter was satisfied that the subject understood the task. This short pilot run was then terminated, usually within 15 trials, and the experiment was started after the experimenter left the chamber. Each subject completed a total of 13 runs. Each run consisted of 50 trials in an adaptive two-down one-up design which tracks the subject's 70.7% correct-response threshold (Wetherill and Levitt, 1965; Levitt, 1971). The first two runs, as well as the last run, measured the subjects' threshold for a single dichotic click; this was a control condition to which thresholds from the lag-click conditions were to be compared. In runs 3–12, the stimulus in each interval of a trial consisted of two clicks, the first representing the lead diotic event and the second representing the lag dichotic event. The ICI was 2 ms. This value has been shown to produce a strong precedence effect for impulsive sounds (Wallach *et al.*, 1949; Zurek, 1980; Zurek and Saberi, 2003), and was selected to be near the training value used by Saberi and Perrott (1990).

On the first interval of each trial of the single-click condition, the dichotic click led to one randomly selected ear by an ITD and, in the second interval, it led to the other ear by the same ITD. The subject's task was to determine if the order of leading ITD was left-ear then right-ear or vice versa. Perceptually, this would be equivalent to determining if the two intracranial sound images in the two intervals of the trial were heard left then right, or right then left. The subject would then press either a left or a right key to respond (left-key response meant that they perceived the sound orders as right to left). Visual feedback was provided after each trial in two forms. First, the subject was informed if s/he was correct. Second, in an image window on the screen, the adaptive track for the current run was displayed which included the ITD values up to the current trial in a graph (i.e., a plot of the trial number versus ITD value). Subjects were instructed to use this trial-by-trial updated graph and the feedback to try to achieve the lowest possible score. A horizontal baseline at 75 μs was plotted on this graph as a target level, and subjects were instructed to attempt to reach values below this line. The initial value of the total ITD on each run was 1300 μs, i.e., 650 μs in each interval. Two successive correct responses led to a reduction of the total ITD by a stepsize of 0.2 log units until the fourth reversal and 0.05 log units

thereafter, rounded to the nearest 25 μs [e.g., an ITD of 1300 μs, after two successive correct responses, would be reduced to $10^{\log(1300)-0.2}=820$ μs; see Saberi (1995)]. An incorrect response led to an increase in total ITD by the stepsize. The 25-μs resolution was determined by the sampling rate of 40 kHz. The variance added from this minimum stepsize to threshold measurements was negligible. Simulations showed that using a 25-μs stepsize increases the standard deviation of threshold estimates by a maximum of 12.5 μs. Lowest and highest possible values of ITD within an interval were 0 and 650 μs, and if the adaptive track required values outside this range, they were corrected to these floor and ceiling values. In all cases, threshold was estimated as the average of the stimulus values at track reversal points, after the fourth reversal. Usually, five to eight reversals went into the calculation of each threshold value. In runs 3–12, the stimulus within each interval consisted of two clicks, a lead diotic click and a lag dichotic click with an ICI of 2 ms. The ITD of the lead click was always zero, and thus carried no information for performing the task. The ITD of the lag click was varied according to the adaptive rules described above.

B. Results

Figure 1 shows results from experiment I. Data from 89 subjects are shown, with 30, 29, and 30 subjects for each row of panels from top to bottom, respectively. Each row shows data for one stimulus intensity. Left panels show threshold interaural delays for a single click, and the right panels are thresholds for the lag click in a two-click design. Each histogram in the right panels represents approximately 300 threshold estimates, i.e., 10 per subject, and each histogram on the left is based on approximately 90 threshold estimates, i.e., 3 per subject. The bin width is 100 μs. Arrows show mean threshold within each panel.

The average population thresholds for the three stimulus intensities of 73, 58, and 43 dB are 247, 275, and 286 μs, respectively, for the single-click condition, and 535, 822, and 838 μs for the lag-click condition. These threshold values from untrained subjects are quite high, but also vary a great deal both across and within subjects as will be discussed in a later section. The histogram modes for the single-click condition are somewhat smaller, being 150 μs for all three intensities. For the lag condition, however, the modes are 1050 μs for the two lower intensities, and a considerably smaller value of 150 μs for the 73-dB intensity. The upper mode of 1050 μs represents the limits imposed on the adaptive track. While the maximum ITD was 1300 μs, the effective ceiling was 1050 μs. Simulations showed that the occurrence of two successive correct responses by chance on a proportion of trials will lead to an effective ceiling of approximately 1050 μs, even if the cue is undetectable.

There is a striking difference between the effects of intensity on single- and double-click conditions. The maximal percentage increase in mean threshold across intensities is 16% for the single-click condition, and 57% for the lag-click condition. Even greater is the difference between the modes of the distribution of thresholds. For the single-click condition, there is a no shift in the mode, while for the lag-click condition there is a 700% decrease from the lowest to the highest intensities and virtually all of this shift occurs between the two highest intensities. Consistent with these observations, Goverts *et al.* (2000) have shown that the precedence effect is most effective for mid-range stimulus intensities.

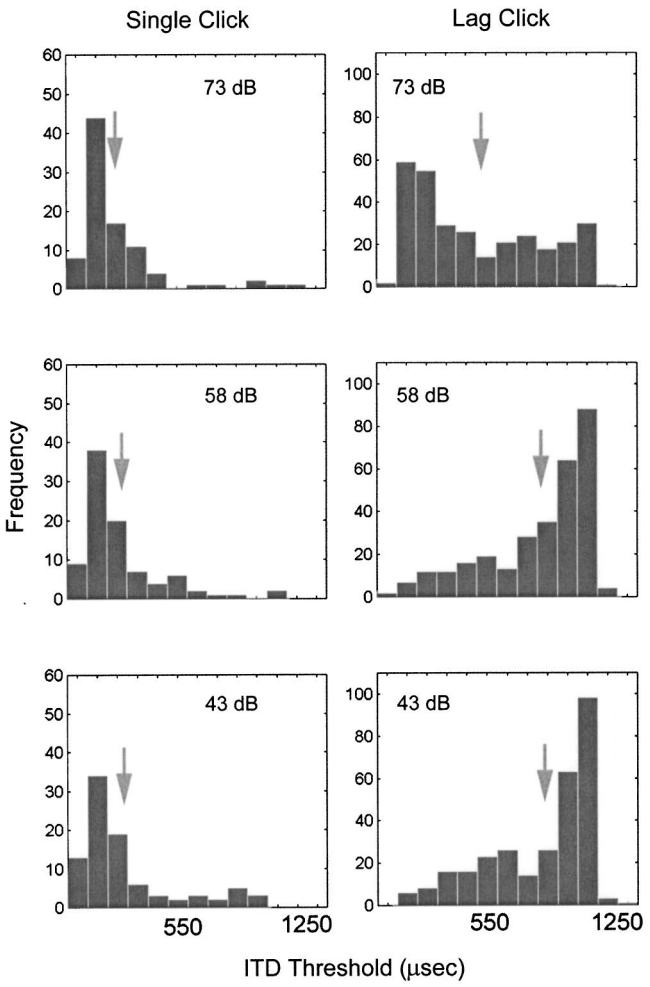


FIG. 1. Results from experiment I. Histograms of interaural delay thresholds in single- and lag-click conditions as a function of stimulus intensity. Arrows show mean thresholds. Data are from 89 subjects.

III. EXPERIMENT II: EFFECTS OF INTERCLICK INTERVAL

A. Method

All methods in this experiment were the same as in experiment I, except that the ICI was a parameter of study and a single threshold at each value of ICI was obtained from each subject. The ICIs were 0.3, 1.0, 1.25, 1.75, 2, 2.35, 5, and 10 ms. Within a run, the value of ICI was fixed. Only two stimulus intensities, 73 and 58 dB, were examined. For the 73-dB condition, thresholds were obtained from 20 subjects, and for the 58-dB condition, from 18 subjects. The subjects differed from those used in experiment I. Data were also obtained from these subjects in a single-click control condition, for which two threshold estimates were obtained and averaged. The single-click conditions were run at the beginning and end of the session. The order in which each subject ran each of the ICI conditions was randomized.

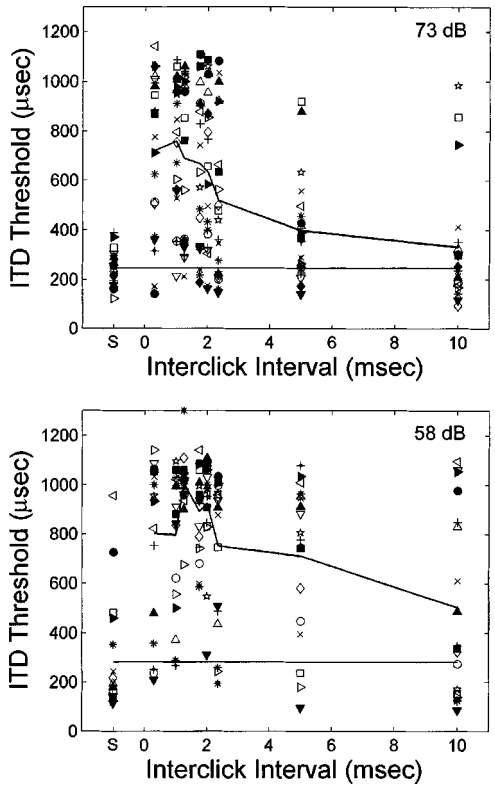


FIG. 2. Results from experiment II. ITD thresholds as a function of interclick interval. Each symbol represents data from one subject, and the solid curve is the mean threshold. The single-click thresholds are plotted on the left of each graph, and the mean single-click threshold is indicated by the horizontal line extending across the graph to facilitate visual comparison. Top and bottom panels shows data from 20 and 18 subjects, respectively, measured at two stimulus intensities.

B. Results

Figure 2 shows results from experiment II. The top panel shows ITD thresholds from individual subjects (different symbols) as a function of ICI at a stimulus intensity of 73 dB, and the lower panel shows thresholds for an intensity of 58 dB. The symbols on the left in each panel show data for the single-click control, and the mean single-click threshold is indicated by the horizontal line extending across the graph to facilitate visual comparison. No significant difference was observed between the two single-click runs at the beginning and end of the run [Wilcoxon nonparametric matched-pairs test $W(18)=56$, ns for the 58-dB intensity and $W(20)=62$, ns for the 73-dB intensity]. The upper curve in each panel shows the mean threshold across all subjects. Note the wide range of thresholds across subjects. The lower panel shows that thresholds for the 58-dB condition are generally similar in form, but higher in value, relative to the 73-dB condition. Thresholds, as expected, are a nonmonotonic function of ICI with a lower mean threshold for the ICI of 0.3 ms compared to the peak of the function at an ICI of 1. The reduced strength of precedence at very low ICIs is typically referred to as ''summing localization'' (Leakey, 1957; Blauert, 1997).

IV. EXPERIMENT III: COMPARISON TO THREHSOLD FROM EXPERIENCED SUBJECTS

In this section we compare thresholds from the population of untrained subjects to those from highly experienced subjects. Previous work has shown that rigorous training leads to improvements in detection of an ITD in the lag click of a two-transient stimulus (Saberi and Perrott, 1990). Others, however, have reported that training has no effect on ITD thresholds in a precedence-effect task (Litovsky *et al.*, 2000). Here we report thresholds for three experienced subjects and show long-term training effects for one subject. The purpose was not an extensive study of training, but to gain better insight into why some subjects may produce significantly lower thresholds than others.

A. Method

Three experienced subjects were used, two having over 10 h (S_e1,S_e2) and the other (S_e3) 50 h of training on precedence-effect tasks. All three had over 100 h of training on other lateralization and localization tasks prior to the start of this experiment. Data from a fourth subject (S_e4) with 66 h of training on the precedence effect will also be described separately. Subject S_e1 was a 20-year-old male, S_e2 was a 25-year-old female, and S_e3 was a 37-year-old male. S_e3 was one of the subjects used in Saberi and Perrott (1990) and is the first author. Subject S_e4 was a completely naïve 21-year-old female listener who was selected for training because her thresholds from experiment I were among the highest of all subjects tested. The stimulus level was 73 dB. The task and stimuli were the same as those described for experiment II. As further controls, the ICI was jittered from interval-to-interval by a random value of up to 10% to distort possible monaural pitch cues, and, in addition, no feedback was provided as to the correctness of the responses. The paradigm used in the first two experiments is referred to as a commutative design. Here, in addition to the commutative design, a second paradigm referred to as a center-side paradigm is used. This paradigm is the same as that employed by Saberi and Perrott (1990) in which both clicks in the first interval of a trial were diotic. In the second interval, the first click was diotic and the second click had an ITD to be detected, either leading to the left or right ear.

B. Results

The upper curves in both panels of Fig. 3 show the curve replotted from the upper panel of Fig. 2 (untrained subjects, commutative task). Error bars represent ± 1 standard error of the mean. The lower curves in the top panel show thresholds from two experienced subjects (S_e1 and S_e3) in the commutative design. These curves represent single-run threshold ITDs in the lag-click condition as a function of ICI. The symbols to the left in each panel show ITD thresholds for the single-click control. Thresholds from experienced subjects, measured using the commutative paradigm, have a peak of about 150 μs, slightly higher than values reported by Saberi and Perrott (1990) who employed a center-side paradigm. We replicated the center-side paradigm, and these results, from three experienced subjects (S_e1, S_e2, and S_e3), are shown in the bottom panel of Fig. 3.

Thresholds obtained from experienced subjects in the center-side paradigm were consistently below 100 μs at all ICIs, are more similar in magnitude to the reported thresholds in Saberi and Perrott (1990), and are significantly

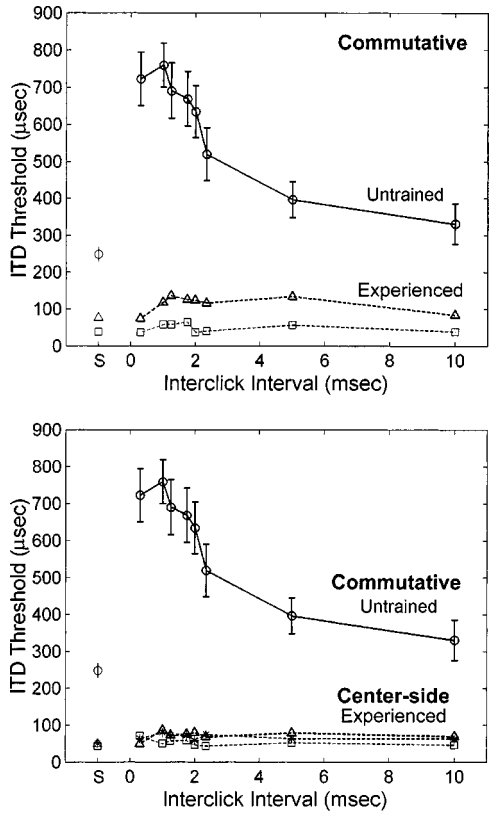


FIG. 3. Lower curves in each panel show thresholds from experienced observers, and the upper curve is the population mean replotted from Fig. 2. Error bars represent ±1 standard error of the mean. The symbols to the left of each panel show thresholds for single-click conditions. Top and bottom panels show data collected from experienced subjects using the commutative and center-side designs, respectively (see text). Data from untrained subjects in both panels were collected using the commutative design.

smaller than those reported by Litovsky *et al.* (2000) who also used the center-side design. It is interesting that the center-side design has produced smaller thresholds compared to the commutative design. If one assumes that the presence of an interaural disparity causes an increase in the variance of the interaural cues within the internal composite stimulus, and not always a consistent lateralization, then it may be easier to compare a stimulus that contains an interaural disparity to a diotic stimulus than to compare two stimuli with interaural delays.[2]

A detailed examination of the adaptive tracks for experienced and untrained subjects revealed unexpected and informative patterns. Panel (a) of Fig. 4 shows thresholds from one experienced subject (S_e1) in the center-side paradigm. During this single run of threshold estimates as a function of ICI, this subject showed very low thresholds at all ICIs, except at an ICI of 1.25 ms, for which the threshold was over 300 μs (upper asterisk). When we inquired, this subject reported that he perceived ''reversals'' during that particular run, in that the side to which the stimulus ITD led was opposite to what he perceived on a subset of trials (the interaural delay became larger even though the subject was convinced that he was responding correctly). The subject repeated this single condition on the very next run, and showed threshold improvements of 459% (lower asterisk), from a value of 340 to 74 μs. This improvement occurred within one run.[3] We inquired if the subject had been inattentive during the high-threshold run, and the subject was insistent that he was fully attentive and perceived a reversed cue.

Panel (b) of Fig. 4 shows the adaptive track for the high-threshold run by this subject [upper asterisk in (a)], and panel (c) shows the adaptive track for the low-threshold run [lower asterisk in (a)]. These tracks are instructive in that they show

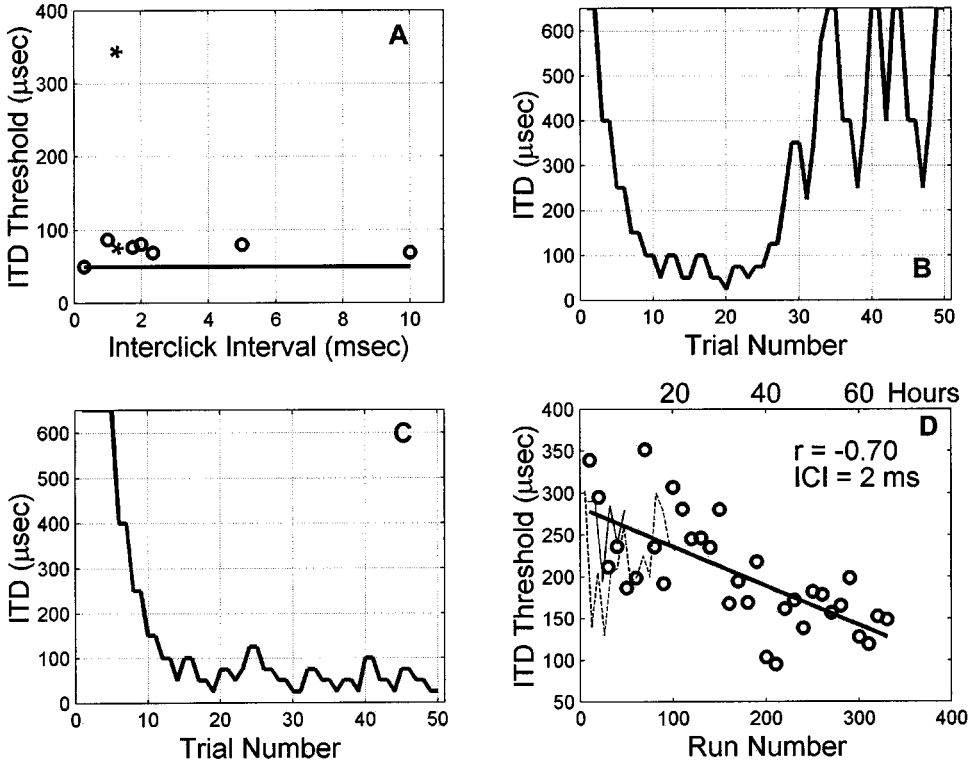


FIG. 4. (a) data from one experienced subject (S_e1). Each circle represents one threshold run in the center-side design. The asterisks are thresholds from two consecutive runs at the same ICI of 1.25 ms. A 459% improvement in threshold is observed within one run. (b) Adaptive track from the upper asterisk in (a). (c) Adaptive track from lower asterisk in (a). (d) Training data from subject S_e4 grouped into ten runs per point. The lower abscissa shows the number of runs and the upper abscissa shows the number of hours of training. The fitted line is a linear regression. The dashed curve and solid curve are data of subjects S3 and S7 from Litovsky *et al.* (2000).

that the high threshold is a result of averaging very low values of the stimulus with very high values. This pattern is unlike what is expected from a subject with a high sensory threshold and implies a dual process: at the early stages, the ITD cue is detectable whereas at the late stages of the track, consistent with the subject's report, an apparent reversal of the percept or possibly a loss of the signal cue is suggested. On the next run at this ICI, the track shows a consistently low trajectory [panel (c)]. As will be discussed later, tracks such as that shown in panel (b), which were numerous for the untrained subjects, suggest a nonstationary discrimination process.

In general, the intersubject variability was quite high. Some subjects showed moderately low thresholds from the start and others showed quite high thresholds. The experienced subjects had prior extensive experience in lateralization tasks and had shown low thresholds from the beginning of the current experiment. It is therefore difficult, at least for two of the experienced subjects, to determine whether their low lag-click thresholds are a result of extensive prior exposure to lateralization and precedence tasks or having *a priori* low-threshold.

For comparison, we selected one of the poorest performing subjects from the pool of all untrained subjects. We then trained this subject at an ICI of 2 ms for several months. In addition to an hourly wage, the subject was given a financial reward whenever the averaged threshold for a 2-h session was lower than all previous sessions. The training data for this subject is shown in panel (d) of Fig. 4, grouped into runs of 10. After 330 runs (~66 h) in the commutative design, it is apparent that the subject's thresholds had improved to among the best of the group. The subject could not continue the training because of reasons unrelated to the experiment. It is evident that even after 66 h of training, the average thresholds for this subject had not yet reached a lower asymptote. If the performance of other experienced subjects is an indication, this subject's final threshold estimates would have likely been even lower if measured in a center-side design. For comparison, we have also plotted training data from subjects S3 and S7 of Litovsky *et al.* (2000), who produced high thresholds from the start, were inexperienced in lateralization tasks, and maintained a high threshold level for 10 to 20 h [thin dashed and solid lines in Fig. 4(d)]. It would be difficult to see an improvement for subject S_e4 from the current study if the data are limited to initial 20 h of training.

V. EVIDENCE FOR UNSTABLE TRACK TRAJECTORIES

The track shown in panel (b) of Fig. 4 prompted us to examine the adaptive tracks from the population of untrained subjects. Such an abrupt change in track trajectory, seen only for lag-click conditions, is indicative of an unstable psychometric function that is, possibly, a composite of a dual process. We speculate that one process is based on a low sensory threshold, and a second process is based on either a cue reversal or loss of the primary signal cue.

An inspection of tracks from untrained subjects showed that on a significant proportion of runs (~20%) that used short ICIs (<5 ms) irregular patterns uncharacteristic of a single stable cue were observed.[4] Such irregular patterns were rarer or nonexistent for the three experienced subjects, with the only instance being that shown in Fig. 4(b). Six irregular tracks from six untrained subjects, grouped into two types of patterns, are shown in Fig. 5. Other interesting patterns also existed that are not shown, such as two-peak or single-trough patterns. The left panels of Fig. 5 show U-shaped trajectories which, until near the end of the run, are typical of a low-threshold run, and are elevated at the end. The right panels show performance consistent with a low threshold at the beginning and end of the track, and an abrupt elevation of threshold at the middle part. The horizontal line within each panel is the 70.7% threshold for that run.

VI. DISCUSSION

The current study investigated lateralization performance in a precedence effect task for a population of subjects with no prior experience in studies of sound localization. Because previous studies have reported large disparities in thresholds (Saberi and Perrott, 1990; Litovsky *et al.*, 2000; Gaskell, 1983; Zurek, 1987; Perrott *et al.*, 1989; Shinn-Cunningham *et al.*, 1993; Yost, 1984) baseline data were collected from large groups of untrained individuals. Population ITD thresholds for single- and lag-click conditions showed high intersubject variability and dependence on intensity.

A comparison of within- to across-subject variability revealed differences between categories of subjects, in that some subjects from the outset displayed considerably lower thresholds than others. A small number of naïve listeners consistently produced low lag-click thresholds. Others produced generally low thresholds interspersed with one or two high values, and still another group consistently produced high thresholds (>400 μs) similar to some early studies (Zurek, 1980; Gaskell, 1983). Individuals who were experienced in lateralization showed low lag-click thresholds less than 100 μs at all ICI in the center/side task and less than 150 μs in the commutative design. In addition, one subject who was not experienced in lateralization and who showed variable thresholds that on average were among the highest of all subjects, during months of training, showed a gradual decline in ITD threshold in the commutative task, suggestive of a slow learning process for this subject.

Several factors may have contributed to the observed low thresholds for some subjects. First, a subject's previous exposure to lateralization tasks in general, including precedence-effect tasks, may be important. Second, some subjects from the start display low pretraining thresholds, whether experienced or not. Third, data from one subject suggests that some subjects with high initial thresholds may improve with extensive training. Fourth, low lag-click thresholds may also be related to stimulus features such as intensity, or to experimental design factors such as financial incentive for improved performance.

Wright and Fitzgerald (2001) have recently reported data on learning in binaural tasks that may have bearing on the present study. They have distinguished between two learning processes: a rapid learning effect that occurs within

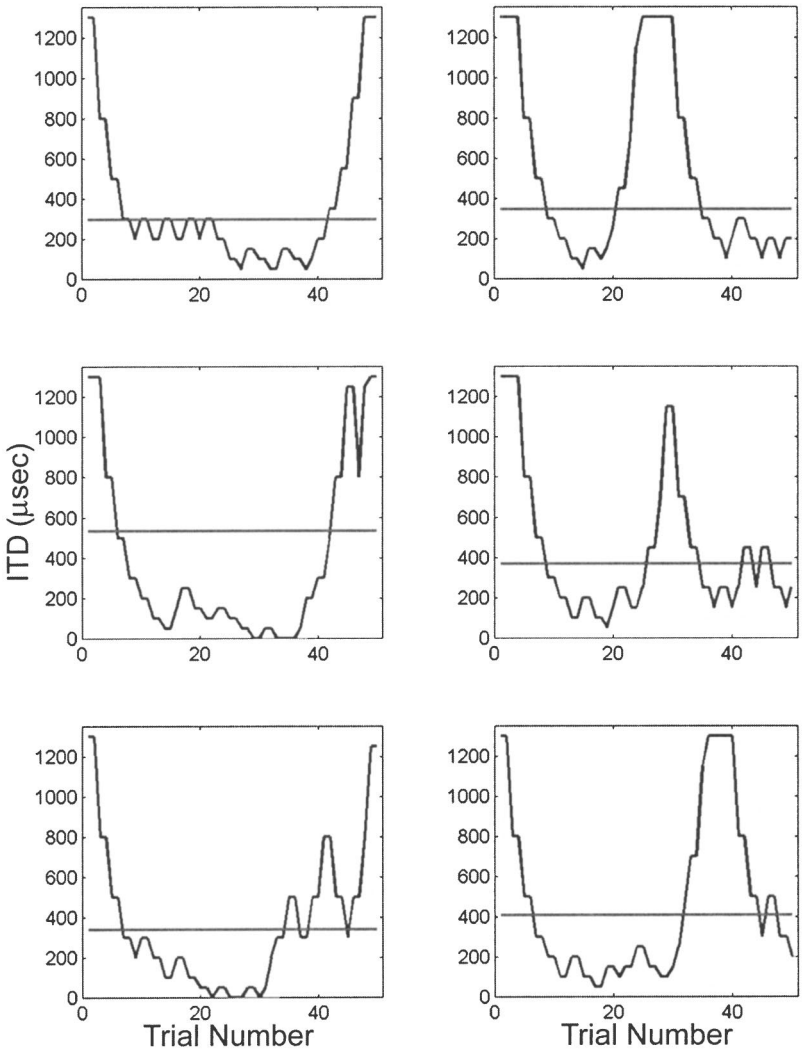


FIG. 5. Sample adaptive tracks from six untrained subjects indicative of a nonstationary lag-click threshold (see text). Solid line shows estimated ITD threshold from the track.

the first few sessions and attributed to procedural learning, and a more gradual learning process attributed to fundamental changes in stimulus processing. The former effect was observed for both ITD and ILD cues, but the latter only for ILD processing. This finding appears different than ours [Fig. 4(d)] and Hafter and Carrier (1970) who show long-term learning for tasks employing ITDs, but similar to others who have reported no long-term learning with low-frequency masking-level-difference or simple ITD discrimination tasks (Bernstein *et al.*, 1998). A close inspection of the individual subject data of Wright and Fitzgerald (2001) shows that although mean thresholds across subjects does not support long-term changes in ITD thresholds, some subjects did display either a long-term gradual decrease in ITD thresholds (their Fig. 1, subjects L9, L10) or a markedly lower post-training threshold (L12, L14). This suggests that even in a simple ITD-discrimination task, intersubject variability in learning may exist (also see subjects L7 and L8 who do not display long-term learning in an ILD-discrimination task). Given the intersubject variability observed by Wright and Fitzgerald, it may be premature to conclude that cue-specific learning in binaural tasks is restricted to one binaural cue and not another.

The large intersubject variability in thresholds reported for untrained subjects in the current study merits further consideration. Figure 6 shows individual thresholds for 30 subjects corresponding to the upper panels of Fig. 1, in addition to two experienced subjects (S_e1 and S_e3, plotted to the right of the dashed lines) who ran under those same conditions (ICI = 2 ms, 73-dB intensity condition). Top and bottom panels of Fig. 6 show lag- and single-click ITD thresholds, respectively. The abscissa shows subject number and the ordinate represents ITD threshold. Each symbol represents one threshold estimate: ten per subject in the top panel and three per subject in the bottom panel. The asterisks in each panel show the threshold estimates for the last run of each condition. These are the 12th and 13th runs of the experiment for the lag- and single-click conditions, respectively. Note that the last run for each condition and subject sometimes produces the lowest, sometimes the highest, and sometimes middle values of thresholds. Note, in addition, that for the lag-click condition, thresholds span the entire range of interaural delays.

A close inspection, however, reveals across subject variability in overall performance. The arrows shown on the bottom axis of the top panel specify subjects who produced relatively low thresholds across most of their runs. Some of these subjects, however, did produce one or two high-

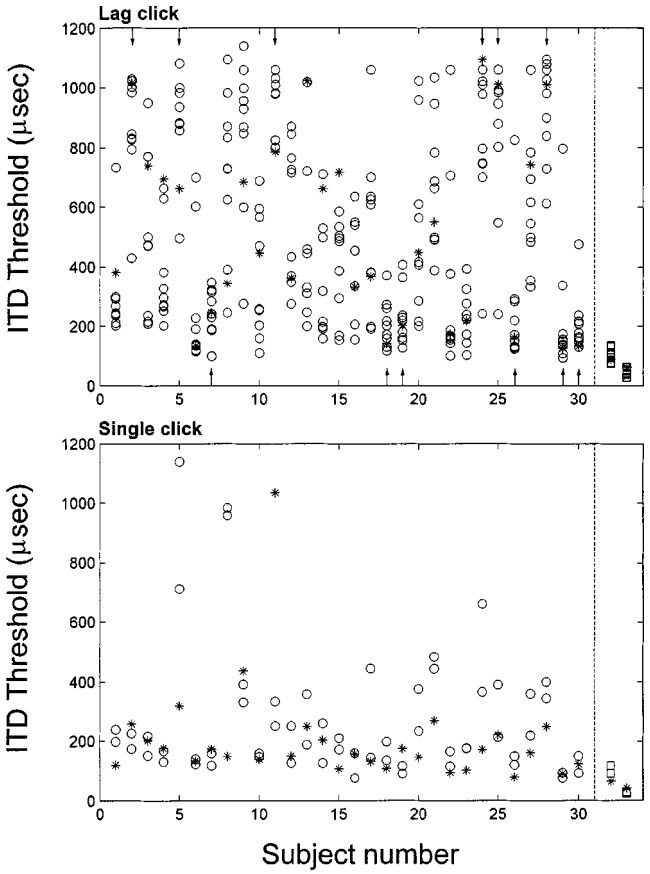


FIG. 6. Single-run thresholds for 30 subjects from the lag- and single-click conditions (top and bottom panels, respectively). Data from untrained subjects correspond to the upper panels of Fig. 1. The ICI was 2 ms and the stimulus intensity was 73 dB. Data from two experienced subjects are also shown to the right of the dashed lines. Each circle or asterisk represents one threshold estimate. The asterisks are the final run for each subject at each condition (lag or single click). Arrows on the lower axis of the top panel specify subjects with relatively low thresholds, and the arrows on the top axis specify subjects with high thresholds.

threshold runs. For example, subjects 26 and 30 show a single outlier run. There seemed to be no set pattern to the order of occurrence of these outlier runs (i.e., beginning or end of the experimental run). Other subjects whose thresholds were clustered at low ITD values produced two or three high-threshold outliers (e.g., subjects 6, 22, and 29). The single-click thresholds for these subjects are consistently low. Other subjects specified by the arrows on the top axis of the top panel show predominately high lag-click thresholds, and other subjects show highly variable thresholds. For these latter subjects, single-click thresholds are reasonably low, although for a very few cases, thresholds are high or variable. Other studies have also shown large intersubject variability in precedence-effect tasks. For example, thresholds for subjects S7 and S8 from Litovsky *et al.* (2000) have averaged lag-click thresholds that differ by a factor of 10 (~20 vs. 200 μs at an ICI of 2.35 ms).

Thresholds for the two experienced subjects shown in Fig. 6 are consistently low. Although some untrained subjects produce mean lag-click thresholds that approach those of experienced subjects, the average is at least twice as large and show greater within-subject variance. For example, the lowest mean threshold for the 30 untrained subjects in Fig. 6 is produced by subject 18 (205 μs). Subjects 29 and 30 also appear to produce generally low thresholds. Excluding the two high outliers for subject 29 and one outlier for subject 30, their thresholds remain higher than those obtained from the two experienced subjects whose mean thresholds are 103 and 49 μs, respectively. The experienced subjects also produced low single-click thresholds as shown in the bottom panel of Fig. 6.

One explanation for outlier thresholds and/or nonmonotonic adaptive tracks reported here may be the pattern of cross-correlation activity generated by dual-impulse stimuli. These patterns are often complex and may result in ambiguous position cues (Zurek and Saberi, 2003; Saberi and Perrott, 1995). Others have also shown that a number of observations related to the precedence effect may be accounted for by examining the complex pattern of cross-correlation activity resulting from dual-pulse stimuli (Saberi and Perrott, 1995; Tollin and Henning, 1998, 1999; Hartung and Trahiotis, 2001; Zurek and Saberi, 2003). The loss of the signal cue may also be related to adaptation (Hafter *et al.*, 1988), or modulation of attention or other cognitive factors (Clifton, 1987; Hartmann and Rakerd, 1989; Blauert and Col, 1991).

Finally, we also observed an effect of stimulus intensity on lag-click thresholds, consistent with Goverts *et al.* (2000) who showed that the precedence effect is most effective at mid-range stimulus levels. The reduced strength of precedence at high stimulus levels may be related to level-dependent neurophysiological effects such as the widening of the frequency tuning of peripheral auditory filters, saturation of neuronal responses, or changes in phase-frequency response of auditory nerve fibers (Allen, 1983; Ruggero *et al.*, 1992). Goverts *et al.* have reported that for sensation levels below about 40 and above 50 dB, the strength of the precedence effect is reduced. A similar effect was observed when the stimulus level was held constant and background noise level was increased (Chiang and Freyman, 1998; Goverts *et al.*, 2000). Interestingly, individuals with mild sensory neural hearing loss also show a decline in the strength of the precedence effect (Goverts *et al.*, 2002). At low sensation levels, internal neural noise may affect onset dominance in the same manner as increasing the level of background external noise. The cause of this weakening of the precedence effect when the signal level is close to the noise floor (internal or external) is not clear.

In summary, findings from the current study show that (1) thresholds from untrained subjects displayed large intersubject variability, (2) the best untrained subjects produced mean thresholds near, but not as low as, those obtained from experienced subjects, (3) the one untrained high-threshold subject tested improved with training over a long period of testing, (4) the center-side stimulus design produced lower thresholds than a commutative design, and (5) poor performance may result from uncertainty with regard to the cue for which to listen. Adaptive tracks for untrained subjects were often unstable, consistent with either losing the cue or listening to the wrong cue in an ambiguous stimulus.

ACKNOWLEDGMENTS

This work was supported by a Grant from the National Institutes of Health (DC-03648). We thank Dr. David R. Perrott, Dr. Pat Zurek, Agavni G. B. Pertosyan, Kyle Nakamoto, and Prisilia Tirtabudi for helpful comments.

[1]For comparison, a continuous 10-Hz train of clicks had intensities of 51, 66, and 81 dB, and at 100-Hz these intensities were 60, 75, and 90 dB.
[2]This idea was suggested by the Associate Editor, Dr. Leslie Bernstein.
[3]Note that the lower threshold was plotted in Fig. 3.
[4]Nearly all naïve listeners showed such patterns and the 20% refers to the approximate proportion of all runs. These were visually categorized by experimenters as the track types described in the text (e.g., U-shaped, single or double peaks, or a sharp trough type).

Allen, J. B. (**1983**). ''Magnitude and phase-frequency response to single tones in the auditory nerve,'' J. Acoust. Soc. Am. **73**, 2071–2092.
Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**). ''Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise,'' J. Acoust. Soc. Am. **103**, 2069–2078.
Blauert, J. (**1971**). ''Localization and the law of the first wavefront in the median plane,'' J. Acoust. Soc. Am. **50**, 466–470.
Blauert, J. (**1989**). ''Binaural technology: Fundamentals and applications,'' J. Acoust. Soc. Am. Suppl. 1 **86**, S66.
Blauert, J. (**1997**). *Spatial Hearing* (MIT, Cambridge, MA).
Blauert, J., and Col, J. (**1991**). ''Irregularities in the precedence effect,'' in *Auditory Physiology and Perception: Proceedings of the 9th International Symposium on Hearing*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 531–538.
Chiang, Y. C., and Freyman, R. L. (**1998**). ''The influence of broadband noise on the precedence effect,'' J. Acoust. Soc. Am. **104**, 3039–3047.
Clifton, R. K. (**1987**). ''Breakdown of echo suppression in the precedence effect,'' J. Acoust. Soc. Am. **82**, 1834–1835.
Clifton, R. K., and Freyman, R. L. (**1989**). ''Effect of click rate and delay on breakdown of the precedence effect,'' Percept. Psychophys. **46**, 139–145.
Cranford, J., and Oberholtzer, M. (**1976**). ''Role of neocortex in binaural hearing in the cat: II. The 'precedence effect' in sound localization,'' Brain Res. **111**, 225–239.
Djelani, T., and Blauert, J. (**2001**). ''Investigations into the build-up and breakdown of the precedence effect,'' Acustica **87**, 253–261.
Franssen, N. V. (**1960**). ''Some considerations on the mechanism of directional hearing,'' Ph.D. thesis, Technische Hogeschool, Delft, The Netherlands.
Freyman, R. L., Zurek, P. M., Balakrishnan, U., and Chiang, Y.-C. (**1997**). ''Onset dominance in lateralization,'' J. Acoust. Soc. Am. **101**, 1649–1659.
Gardner, M. B. (**1968**). ''Historical background of the Haas and/or precedence effect,'' J. Acoust. Soc. Am. **43**, 1243–1248.
Gaskell, H. (**1983**). ''The precedence effect,'' Hear. Res. **12**, 277–303.
Goverts, S. T., Houtgast, T., and van Beek, H. H. M. (**2000**). ''The precedence effect for lateralization at low sensation levels,'' Hear. Res. **148**, 88–94.
Goverts, S. T., Houtgast, T., and van Beek, H. H. M. (**2002**). ''The precedence effect for lateralization for the mild sensory neural hearing impaired,'' Hear. Res. **163**, 82–92.
Haas, H. (**1949**). ''The influence of a single echo on the audibility of speech,'' J. Audiol. Eng. Soc. **20**, 145–159, English translation (1972).
Hafter, E. R., and Carrier, S. C. (**1970**). ''Masking level differences obtained with pulsed tonal maskers,'' J. Acoust. Soc. Am. **47**, 1041–1047.
Hafter, E. R., and Dye, R. H. (**1983**). '' Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 1708–1713.
Hafter, E. R., Buell, T. N., and Richards, V. M. (**1988**). ''Onset-coding in lateralization: Its form, site, and function,'' in *Auditory Function*, edited by G. M. Edelman, W. E. Gail, and W. M. Cowan (Wiley, New York).
Hafter, E. R., Dye, R. H., and Wenzel, E. (**1983**). ''Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 644–651.
Hartmann, W. M., and Rakerd, B. (**1989**). ''Localization of sound in rooms. 4. The Franssen effect,'' J. Acoust. Soc. Am. **86**, 1366–1373.
Hartung, K., and Trahiotis, C. (**2001**). ''Peripheral auditory processing and investigations of the 'precedence effect' which utilize successive transient stimuli,'' J. Acoust. Soc. Am. **110**, 1505–1513.
Hays, W. L. (**1981**). *Statistics*, 2nd ed. (CBS College, New York).
Hochster, M. E., and Kelly, J. B. (**1981**). ''The precedence effect and sound localization by children with temporal lobe epilepsy,'' Neuropsychologia **19**, 49–55.
Leakey, D. M. (**1957**). ''Some measures on the effects of interchannel intensity and time differences in two channel sound systems,'' J. Acoust. Soc. Am. **31**, 977–986.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lindemann, W. (**1986a**). ''Extension of a binaural cross-correlation model by contralateral inhibition: I. Simulation of lateralization for stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622.
Lindemann, W. (**1986b**). ''Extension of a binaural cross-correlation model by contralateral inhibition: II. The law of the first wavefront,'' J. Acoust. Soc. Am. **80**, 1608–1622.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.
Litovsky, R. Y., Hawley, M. L., Fligor, B. J., and Zurek, P. M. (**2000**). ''Failure to unlearn the precedence effect,'' J. Acoust. Soc. Am. **108**, 2345–2352.
Mickey, B. J., and Middlebrooks, J. C. (**2001**). ''Responses of auditory cortical neurons to pairs of sounds: Correlates of fusion and localization,'' J. Neurophysiol. **86**, 1333–1350.
Muncey, R. W., Nickson, A. F. B., and Dubout, P. (**1953**). ''The acceptability of speech and music with a single artificial echo,'' Acustica **3**, 168–173.
Perrott, D. R., Marlborough, K., Merrill, P., and Strybel, T. Z. (**1989**). ''Minimum audible angel thresholds obtained under conditions in which the precedence effect is assumed to operate,'' J. Acoust. Soc. Am. **85**, 282–288.
Pitman, J. (**1993**). *Probability* (Springer Verlag, New York).
Rakerd, B., Hartmann, W. M., and Hsu, J. (**2000**). ''Echo suppression in the horizontal and median sagittal planes,'' J. Acoust. Soc. Am. **107**, 1061–1064.
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. 85–92.
Saberi, K. (**1995**). ''Some considerations on the use of adaptive methods for estimating interaural-delay thresholds,'' J. Acoust. Soc. Am. **98**, 1803–1806.
Saberi, K. (**1996**). ''Observer weighting of interaural delays in filtered impulses,'' Percept. Psychophys. **58**, 1037–1046.
Saberi, K., and Perrott, D. R. (**1990**). ''Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate,'' J. Acoust. Soc. Am. **87**, 1732–1737.
Saberi, K., and Perrott, D. R. (**1995**). ''Lateralization of click-trains with opposing onset and ongoing interaural delays,'' Acustica **81**, 272–275.
Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). ''Adjustment and discrimination measurements of the precedence effect,'' J. Acoust. Soc. Am. **93**, 2923–2032.
Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (**1995**). ''Cross-frequency interactions in the precedence effect,'' J. Acoust. Soc. Am. **98**, 164–171.
Tollin, D. J., and Henning, G. B. (**1998**). ''Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect,'' J. Acoust. Soc. Am. **104**, 3030–3038.
Tollin, D. J., and Henning, G. B. (**1999**). ''Some aspects of the lateralization of echoed sound in man. II. The role of stimulus spectrum,'' J. Acoust. Soc. Am. **105**, 838–849.
Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **52**, 315–336.
Wetherill, G. B., and Levitt, H. (**1965**). ''Sequential estimation of points on a psychometric function,'' Br. J. Math. Stat. Psychol. **18**, 1–10.
Wright, B. A., and Fitzgerald, M. B. (**2001**). ''Different patterns of human discrimination learning for two interaural cues to sound-source location,'' Proc. Natl. Acad. Sci. U.S.A. **98**, 12307–12312.
Yin, T. C. T., and Litovsky, R. Y. (**1995**). ''Physiological studies of the precedence effect in the inferior colliculus of the cat,'' in *Advances in*

Hearing Research, edited by G. A. Manley, G. M. Klump, C. Koppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore).
Yost, W. A., and Soderquist, D. R. (**1984**). ''The precedence effect—revisited,'' J. Acoust. Soc. Am. **76**, 1377–1383.
Zurek, P. M. (**1980**). ''The precedence effect and its possible role in the avoidance of interaural ambiguities,'' J. Acoust. Soc. Am. **67**, 952–964.
Zurek, P. M. (**1987**). ''The precedence effect,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York).
Zurek, P. M., and Saberi, K. (**2003**). ''Lateralization of two-transient stimuli,'' Percept. Psychophys. **65**, 95–106.

Vertical-plane sound localization probed with ripple-spectrum noise[a)]

Ewan A. Macpherson[b)] and John C. Middlebrooks
Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506

(Received 22 July 2002; revised 19 March 2003; accepted 21 April 2003)

Ripple-spectrum stimuli were used to investigate the scale of spectral detail used by listeners in interpreting spectral cues for vertical-plane localization. In three experiments, free-field localization judgments were obtained for 250-ms, 0.6–16-kHz noise bursts with log-ripple spectra that varied in ripple density, peak-to-trough depth, and phase. When ripple density was varied and depth was held constant at 40 dB, listeners' localization error rates increased most (relative to rates for flat-spectrum targets) for densities of 0.5–2 ripples/oct. When depth was varied and density was held constant at 1 ripple/oct, localization accuracy was degraded only for ripple depths $\geqslant$20 dB. When phase was varied and density was held constant at 1 ripple/oct and depth at 40 dB, three of five listeners made errors at consistent locations unrelated to the ripple phase, whereas two listeners made errors at locations systematically modulated by ripple phase. Although the reported upper limit for ripple discrimination is 10 ripples/oct [Supin *et al.*, J. Acoust. Soc. Am. **106**, 2800–2804 (1999)], present results indicate that details finer than 2 ripples/oct or coarser than 0.5 ripples/oct do not strongly influence processing of spectral cues for sound localization. The low spectral-frequency limit suggests that broad-scale spectral variation is discounted, even though components at this scale are among those contributing the most to the shapes of directional transfer functions. [DOI: 10.1121/1.1582174]

PACS numbers: 43.66.Qp, 43.66.Ba [LRB]

I. INTRODUCTION

Localization of sound sources in the vertical plane (i.e., in the vertical and front/back dimensions), relies upon the direction-dependent filtering performed by the pinnae, head, and upper body. This filtering provides spectral cues that complement the binaural difference cues (ITD and ILD), which provide lateral angle (right/left) information. The role of spectral cues in vertical plane localization has been recognized and investigated for over a century (e.g., Pierce, 1901; Strutt, 1907; Burger, 1958; Blauert, 1969/70; Middlebrooks and Green, 1991), but many aspects of the processing of spectral cues to source location remain unclear. These aspects include the nature of the spectral cues themselves (for example, whether whole spectral profiles or the frequency loci of individual spectral features such as notches or peaks are most informative), the scale of spectral detail required or utilized by human listeners, and the ability of the auditory system to differentiate between source-spectrum features and directional cues. In the present study, we examined the latter two aspects of spectral cue processing in a manner intended to be independent of the issue of whether the spectral cues are localized or distributed in frequency.

The filtering to which incident sound from a particular direction is subjected can be described by the *directional transfer function* (DTF), which is the directional component of the head-related (or source-to-eardrum) transfer function (Wightman and Kistler, 1989; Middlebrooks and Green, 1990; Middlebrooks, 1999a). In the frequency domain, the spectra of these transfer functions show location-dependent variation across frequency both at coarse (multioctave) and fine (suboctave) scales. The scale of spectral detail actually important in auditory processing of the DTF-based cues has been explored previously by others by removing spectral detail from the synthesis filters in virtual auditory space localization tasks. This has been accomplished by truncating the DTF impulse responses in the time domain (Zahorik *et al.*, 1995), by removing fine-scale components of the DTF profiles in the spectral-envelope frequency domain (Kulkarni and Colburn, 1998), by flattening the DTFs over various limited frequency ranges (Langendijk and Bronkhorst, 2002), and by modeling the DTFs with low-order autoregressive moving average filters (Asano *et al.*, 1990). The first two of these operations result in wholesale smoothing of the DTF spectra. The third method is a local operation that removes all spectral features smaller than the width of the flattened region, and the fourth also smooths the spectrum nonuniformly across frequency.

Analogous to the representation of a time-domain signal as a sum of sinusoids of various frequencies, a spectral profile may be described as a sum of periodic frequency-domain components that differ in their *spectral-envelope frequency* (measured, for example, in ripples/octave or ripples/kHz). The manipulations described above can all be thought of as various forms of low-pass filtering of the DTFs in this spectral-envelope frequency (or *ripple density*) domain. Thus, all of these manipulations removed information over a rather wide range of spectral-envelope frequencies.

There is a growing body of evidence, both psychophysi-

[a)]Parts of this work were presented previously in poster form at the 139th Meeting of the Acoustical Society of America, Atlanta, GA, June, 2000.
[b)]Electronic mail: emacpher@umich.edu

cal (Bero and Eddins, 2001; Eddins and Bero, 2001; Eddins and Harwell, 2002) and physiological (e.g., Shamma *et al.*, 1995; Versnel and Shamma, 1998), that the auditory system analyzes spectral profiles in a manner similar to the spatial-frequency-channel analysis present in the visual system (e.g., Campbell and Robson, 1968; De Valois *et al.*, 1982). That is, spectral profiles appear to be represented in multiple spectral-envelope frequency channels. Components of the profile with similar spectral-envelope frequencies interact (exhibiting interactions similar to energetic masking, for example), but those with dissimilar spectral-envelope frequencies do not. Like any profiles, DTF spectra can be decomposed into independent ripple components with differing spectral-envelope frequencies, and it is possible that processing of spectral cues occurs in a limited range of envelope frequency channels.

A variety of studies (e.g., Bernstein and Green, 1987; Supin *et al.*, 1999) has indicated that sensitivity to spectral ripples, as measured by absolute detection thresholds or phase-discrimination thresholds, declines sharply at ripple densities above 3 ripples/oct. This accords well with critical bandwidths derived from classical tone-in-noise masking experiments (e.g., Zwicker and Terhardt, 1980). Although sensitivity to ripple phase above 3 ripples/oct is substantially reduced, it is not abolished, and the upper limit of ripple-phase discrimination appears to be approximately 10 ripples/oct at center frequencies above 1–2 kHz (Supin *et al.*, 1999). As we show below, DTFs themselves have most of their ripple energy below 3 ripples/oct, but it does not necessarily follow that the information-bearing features of the spectra fall primarily in this range or that listeners concentrate their processing at this scale of detail. Our experiments were designed to examine this issue.

In the present study, listeners performed a free-field localization task and indicated the apparent direction of ripple- or flat-spectrum targets by orienting with their heads. We proceeded on the assumption, derived from previous studies (e.g., Blauert, 1969/70; Langendijk and Bronkhorst, 2002; Middlebrooks, 1992; Rakerd *et al.*, 1999), that the auditory system is often unable to differentiate between those features of the proximal stimulus (ear canal) spectrum resulting from DTF filtering and those originally present in the distal (source) spectrum. Although listeners can adapt to room acoustics that initially cause localization errors (Shinn-Cunningham and Kopčo, 2002), this appears not to be possible for wideband, irregular source spectra that cause localization errors (Macpherson, 1998). Thus, target-spectrum ripples were expected to obscure the DTF spectra *at specific scales of detail*, to interfere with their extraction and recognition, and to produce consistent errors in vertical-plane localization.

The approach of *adding* spectral variation to target source spectra rather than *removing* information from the DTFs was motivated by two considerations. First, by adding single ripple components to the target source spectra, we potentially disrupted DTF-cue processing within a limited range of spectral-envelope frequencies at one time. Second, manipulation of only the source spectrum ensured that if listeners *were* able to separate the source and DTF components of the stimulus spectrum, accurate spectral localization cues would still be available.

The following sections of this paper describe three psychophysical experiments. In Experiment I, listeners localized free-field, ripple-spectrum targets of various ripple densities at a fixed ripple depth. We examined the rate and locations of large polar-angle (vertical-plane) localization errors as a function of ripple density. In Experiment II, listeners performed the same localization task for targets with a fixed ripple density (1 ripple/oct), but with various ripple depths. We examined the rate and locations of errors as a function of ripple depth. In Experiment III, we presented 1-ripple/oct targets with various ripple phases at a fixed ripple depth and analyzed the rate and locations of errors as a function of phase. We also examined the relations among the phases of the source-spectrum ripple profiles, the listeners' DTF spectra, and the locations of listeners' erroneous localization responses.

II. METHODS

A. Subjects

Six paid listeners (two female: S64 and S67; and four male: S63, S65, S66, and S77) aged 19–26 years were recruited from the University of Michigan community. All listeners had hearing loss $\leqslant$10 dB at octave frequencies from 0.25–8 kHz as determined by standard clinical audiometry. None had previous experience in sound localization or other psychoacoustic experiments. Before data collection began, all were given instruction and several hours of practice in the psychophysical localization procedure using flat-spectrum, broadband noise targets. All six listeners participated in Experiments I and II, and all except S63 participated in Experiment III.

B. Stimulus synthesis

The target stimuli consisted of 250-ms bursts of shaped-spectrum, random-phase noise with 20-ms raised-cosine (i.e., $\cos^2$) onset and offset ramps. Stimulus waveforms were synthesized digitally using the inverse Fourier-transform method. The spectra of the unfiltered control stimuli were flat from 0.6 to 16 kHz with sharp roll-offs ($\geqslant$80 dB/oct) above and below these frequencies and no energy below 0.3 kHz or above 18 kHz. Stimuli were manipulated by imposing log-frequency ripple patterns on the 1–16-kHz portion of the flat spectrum. Expressing frequency, f, in octaves relative to 1000 Hz as $\Omega(f)=\log_2(f/1000)$, the relative spectrum level (in dB) was determined by the expression

$$S_{\mathrm{dB}}(f)=\frac{A}{2}\cdot\sin(2\pi\cdot\rho\cdot\Omega(f)+\phi).$$

The parameters of the ripple spectra were: peak-to-trough amplitude, A, in dB; density, ρ, in ripples/oct; and phase at 1000 Hz, ϕ, in radians. Raised-cosine ramps 1/8-octave wide were applied in the frequency domain to the high- and low-frequency ends of the ripple profile to smooth any spectral discontinuities. Prior to transformation to the time domain, each spectral profile was divided (over the stimulus passband, 0.3–18 kHz) by the magnitude of the

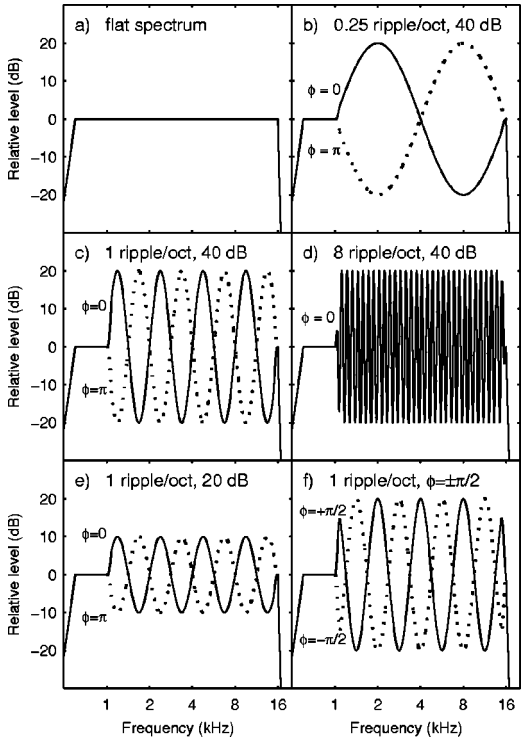


FIG. 1. Examples of ripple spectra with various densities, depths, and phases. (a) flat-spectrum (0.3–16 kHz) control stimulus (Experiments I, II, and III). (b), (c), (d) 0- and π-phase, 40-dB depth ripples of various densities (Experiment I). (e) 1-ripple/oct, 20-dB depth ripples (Experiment II). (f) 1-ripple/oct, 40-dB depth, $\pm\pi/2$-phase ripples (Experiment III).

measured transfer function of the presentation loudspeaker, thus removing the frequency-response characteristics of the loudspeaker from the acoustic stimulus. The method used to measure these transfer functions is described below in Sec. II E.

Examples of some of the ripple spectra used in Experiments I, II, and III are plotted in Fig. 1. In each panel, relative spectrum level in dB is plotted on the ordinate versus frequency (on a logarithmic scale) on the abscissa. Figure 1(a) shows the flat (0.6–16 kHz) spectrum of the control stimuli used in each experiment. Panels (b), (c), and (d) show examples, at densities of 0.25, 1, and 8 ripples/oct, of the 40-dB depth ripple spectra used in Experiment I; solid and dashed lines denote 0 and π phases, respectively. Panel (e) shows examples of the lower-depth, 0- and π-phase, 1-ripple/oct spectra used in Experiment II, and panel (f) shows two examples of spectra with other ripple phases ($\pi/2$ and $-\pi/2$) as used in Experiment III.

C. Stimulus presentation

Stimuli were presented at a sampling rate of 50 kHz using digital-to-analog converters and analog attenuators (Tucker-Davis Technologies [TDT] models DD1 and PA4). The signals were amplified (Adcom GFA-535II) and directed to the appropriate loudspeaker (Pioneer TS-879 two-way coaxial) through a power multiplexer (TDT PM1). Although no explicit reconstruction filter was used, any spectral artifacts lay beyond the 30-kHz bandwidth of the loudspeakers. Stimuli were presented at sensation levels of 45–55 dB from

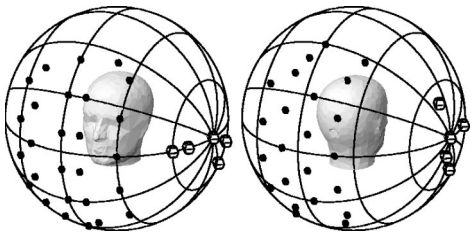


FIG. 2. Front, rear, and lateral target locations used in Experiments I, II, and III. Ripple-spectrum and flat-spectrum noise bursts were presented from front and rear locations on or near the median plane (filled symbols). Only flat-spectrum stimuli were presented from the more lateral locations (open symbols).

one of two loudspeakers mounted opposite each other on a computer-controlled, 1.2-m radius hoop. This hoop could be rotated about horizontal and vertical axes, thus allowing a loudspeaker to be positioned nearly anywhere around a centrally situated listener. The hoop was covered with sound-absorbing foam, and was located in a sound-attenuating anechoic chamber ($2.6\times3.7\times3.2$ m of usable space), the walls and ceiling of which were lined with glass-fiber wedges and the floor with sound-absorbing foam wedges. During the experimental sessions, the anechoic chamber was completely darkened so that the hoop and loudspeakers were not visible. All stimulus generation, hoop movement, and data collection tasks were controlled by custom software written in MATLAB (The Mathworks, Inc.) running on an Intel® based personal computer.

Stimuli were presented in blocks of 110 trials. In each block, bursts with a single set of ripple parameters were interleaved with flat-spectrum, wideband noise targets. One ripple-spectrum and one flat-spectrum trial were presented in a randomized order from each of the 25 front and 25 rear locations illustrated in Fig. 2 (filled symbols). These locations were distributed approximately uniformly in the regions within 30 deg of the vertical median plane and between elevations 60 deg above and below the horizontal plane. Additionally, ten flat-spectrum targets were presented from locations (open symbols) near the interaural axis ($\geqslant$60 deg from the median plane) in order to encourage the listeners to use a full and natural range of lateral angle responses. Each block of trials was completed in $\sim$20 min. Listeners rested after every two blocks, and typically completed four blocks in one 90-min session.

D. Response procedure

Listeners were instructed to indicate the apparent location of a target by turning the body and tilting the head to orient the face in the appropriate direction. Gilkey *et al.* (1995) demonstrated this technique to be more accurate than either a verbal reporting of coordinates or a spherical-model pointing method. The listener stood at the center of the darkened anechoic chamber, and at the beginning of each trial, oriented toward a wideband reference noise source positioned at eye level, 2 m directly in front of the listener. The listener then initiated a trial by pressing a hand-held button. The noise source was silenced, and, after a random delay of 500–1500 ms, a target stimulus was presented from one of

the hoop-mounted loudspeakers. After hearing the stimulus, the listener oriented toward its perceived location, then pressed the button again, which triggered measurement of the listener's head orientation by a head-mounted electromagnetic tracking device (Polhemus FASTRAK). This orientation measurement constituted the listener's response.

Listeners' heads were not restrained during stimulus presentation, but the accuracy of their fixation on the reference noise source before each trial was determined by a head-orientation measurement triggered immediately prior to the stimulus presentation. The root-mean-square values for deviations from the fixation point varied from 2.0 to 3.6 deg among listeners. Both horizontal and vertical deviations were distributed similarly.

E. Directional transfer function measurement

In order to permit comparison of ripple profiles and DTF spectra, we measured each listener's DTFs using the procedure described by Middlebrooks (1999a). Briefly, 512-point, 50-kHz Golay code pairs (Zhou *et al.*, 1992) were presented from a loudspeaker positioned 1.2 m from the listener's head at 400 locations approximately evenly distributed in space around the listener's head, including 32 locations on the median vertical plane. The responses to these excitation signals were recorded simultaneously by two miniature electret microphones (Knowles, model 1934) inserted approximately 5 mm into the listeners' ear canals. This was deep enough to capture all spatial information independent of any ear canal resonances (Hammershøi and Møller, 1996; Middlebrooks *et al.*, 1989).

Head-related transfer functions (HRTFs) were extracted by cross correlation of excitation and response, Fourier transformation, and the removal of the measured loudspeaker transfer function. The loudspeaker transfer functions themselves were obtained similarly, in the absence of the listener, by recording the on-axis response to the Golay code excitation with a 1/2-in. reference microphone (Larson Davis, model 2540) positioned at the location of the listener's head. DTFs for each ear were computed from the set of HRTFs by removing the root-mean-square average magnitude spectrum (i.e., the nondirectional component) from the set of HRTFs for that ear. The nondirectional component also contained the transfer functions of the microphones and any fixed-geometry reflections from the hoop, which were thus removed by this procedure.

III. EXPERIMENT I: EFFECT OF RIPPLE DENSITY

In Experiment I, we measured the influence of ripple density on vertical-plane localization accuracy. We predicted that disruption of localization by source-spectrum ripples should decline at the higher ripple densities in a manner consistent with increasing thresholds for ripple detection and phase discrimination. We also were interested to discover whether sensitivity at lower ripple densities follows the distribution of ripple energy in the DTFs or whether the processing of spectral cues involves some preferred ripple density.

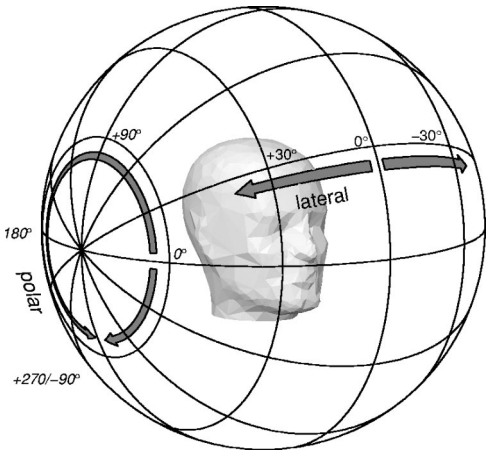


FIG. 3. Horizontal polar coordinate system. Lateral angle is the angle between the location and the median sagittal plane; positive values are to the listener's right. Polar angle combines elevation and front/back position. −90 or +270 deg: below; 0 deg: front; +90 deg: overhead; +180 deg: rear.

A. Stimuli

Target stimuli consisted of ripple-spectrum noise bursts with ripple densities of 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, or 8 ripples/oct, phases of 0 or π, and a peak-to-trough depth of 40 dB. Examples of some of these ripple-spectrum profiles are plotted in Figs. 1(b), (c), and (d). In each block of trials, ripple parameters were held constant and rippled stimuli were interleaved with flat-spectrum targets. The order in which the blocks were completed was selected randomly for each listener. Data were not collected for listener S63 in the 0.75-ripples/oct, 0-phase condition because of his limited availability.

B. Results

1. Coordinate system

Target and response location data were expressed in a horizontal-pole, lateral- and polar-angle coordinate system (Fig. 3). Lateral angle represents the left/right angle between a location and the median plane, and polar angle, the angle of rotation about the interaural axis, captures the elevation and front/rear components of a location.

2. Lateral angle data

We computed measures of bias and scatter in the lateral angle components of each listener's localization responses. This was done separately for the flat- and ripple-spectrum stimuli in each block of trials. The ten flat-spectrum targets that were located near the interaural axis were excluded. The bias measure was the *lateral angle gain* (Hofman and Van Opstal, 1998; Macpherson and Middlebrooks, 2000), i.e., the slope of a linear regression between the target and response lateral angles. The scatter measure was the root-mean-square deviation of the response lateral angles from the values predicted by the regression.

All but one of the listeners (S77) had mean lateral angle gains close to 1 for the flat-spectrum stimuli (range, 0.99–1.12; overall mean, 1.02). The mean flat-spectrum gain for listener S77 was 1.43, indicating consistent overshoot in lat-

eral angle responses. For the ripple-spectrum stimuli, the mean lateral angle gains were slightly higher for all listeners (excluding S77: range, 1.02–1.17; overall mean, 1.11; for S77, 1.53), but the mean increase in gain was $<9\%$.

The mean lateral angle scatter for all listeners was similar for the flat- (range, 10–16 deg; overall mean, 14 deg) and ripple-spectrum (range, 8–16 deg; overall mean, 15 deg) targets. Listener S77 had the highest lateral angle scatter (16 deg for both flat- and ripple-spectrum targets).

These results show that both flat- and ripple-spectrum stimuli were localized similarly in lateral angle, consistent with our finding that spectral cues have little or no contribution to this dimension of apparent position (Macpherson and Middlebrooks, 2002; Middlebrooks, 1992). For that reason, we present only analysis of the polar angle data in the remainder of this paper.

3. Polar angle data

a. Raw polar angle data. The polar angle components of listeners' responses are plotted versus target polar angle in Fig. 4 for five (0.25, 0.75, 1.5, 2, and 6 ripples/oct) of the ten ripple densities presented. (Data for the 1-ripple/oct conditions are plotted with those of Experiment II in Fig. 8.) Each column of panels contains data for one ripple density, and each row contains data for one listener. In each panel, responses for the 0- (open circles) and π-phase ($\times$s) ripple profiles are combined with those for the corresponding flat-spectrum stimuli (small closed circles).

Listeners S63, S64, S65, and S66 localized flat-spectrum stimuli very accurately; slopes of response versus target polar angle (i.e., *polar angle gain*) were close to 1, and very few front/rear reversals were generated. Listeners S67 and S77 had less accurate performance. S67 consistently exhibited an exaggerated polar angle gain for rear hemisphere targets and tended to make rear-to-front reversals for targets in the high-rear quadrant (polar angles 90–180 deg). S77 also made many reversals for rear-hemisphere targets and had slightly exaggerated polar angle gain for front-hemisphere targets. All listener's patterns of responses to flat-spectrum stimuli were consistent from block to block.

A detailed analysis of errors in localizing the ripple-spectrum stimuli follows below, but we note here that greater numbers of deviations from accurate performance occurred for the ripple-spectrum stimuli than for the flat. At a particular ripple density, erroneous responses for the 0- and π-phase spectra tended to form clusters at different locations, which were independent of the target locations. At the 0.25-ripples/oct density, π-phase ripples produced performance similar to that for the flat-spectrum targets, but the 0 phase produced many errors for some listeners. At the highest density shown (6 ripples/oct), the ripple spectra were localized with accuracy similar to that for the flat-spectrum targets.

b. Computation of polar error rate. Our quantitative measure of localization performance for each stimulus spectrum was the proportion of responses exhibiting large errors in the polar angle dimension. ''Errors'' were defined relative to the response to a flat-spectrum stimulus at the target location predicted by a linear fit to the main concentration of responses to flat-spectrum stimuli collected in the same block of trials. As defined, errors primarily reflected deviations from flat-spectrum performance with the effects of spectrally independent motor response bias (e.g., pervasive over- or undershoot) removed. The error-rate measure was computed for both the flat- and ripple-spectrum stimuli in each block of trials.

As an example, we illustrate the computation of the error-rate measure using the polar-angle response data from listener S66 for the trial block containing 1-ripple/octave, 40-dB, π-phase ripple spectra (Fig. 5). *First*, we obtained a pair of linear functions (solid lines) that independently summarized the front- and rear-hemisphere, polar-angle, target-response relations for the flat-spectrum stimuli (squares). Outliers and front/rear reversals were excluded from the computation of the linear functions as explained below, but were included in all further analyses.

In order to exclude outliers and reversals and isolate the main concentration of responses in the computation of the linear fits, an *ad hoc* selective, iterative regression procedure was used (Macpherson and Middlebrooks, 2000). Data for front- (polar angles from -90 to $+90$ deg) and rear- ($+90$ to $+270$ deg) hemisphere targets were analyzed separately. To initialize the procedure, the data set was temporarily limited to trials in which the response fell in the correct (front or rear) hemisphere, and a regression line was fitted to these points. Then, all points lying farther from the line than a criterion distance, Δ, were temporarily eliminated and the regression recomputed using only the remaining data. This procedure was iterated (typically 2–3 times) until it converged. Points eliminated during one iteration were available for reselection in future iterations if the regression line moved closer to them. A value for Δ of 40 deg was found empirically to work well in fitting the regression line to the main concentration of the response distributions while ignoring outliers and reversed responses.

The slopes of these regression lines reflected listeners' typically accurate localization performance for the flat-spectrum stimuli. Across listeners and conditions, the mean slope was 1.04, and means across conditions for each listener varied between 0.80 (S65, front-hemisphere targets) and 1.59 (S67, rear-hemisphere targets). The root-mean-square deviations of those responses within 45 deg of the regression lines were similar to the scatter found for lateral angle (range across listeners, 8–13 deg; overall mean, 10 deg). The proportion of variance accounted for, R^2, was high (range across listeners, 0.85–0.95; overall mean, 0.92).

Next, localization accuracy for flat- and ripple-spectrum targets was assessed by measuring the deviation of responses from the linear predictors. We defined the *polar angle deviation* as the signed deviation of a particular response from the polar angle predicted by the linear fit for the given target location. Responses for which the magnitude of the polar angle deviation was $\leqslant 45$ deg (within the dashed lines) were considered to be *near-veridical* (open symbols), and responses for which this magnitude was >45 deg (beyond the dashed lines) were classified as *polar errors* (filled symbols). The 45-deg criterion was chosen to encompass the scatter observed in the apparently near-veridical populations of flat-spectrum responses (Fig. 4 and Fig. 5) and to match that used

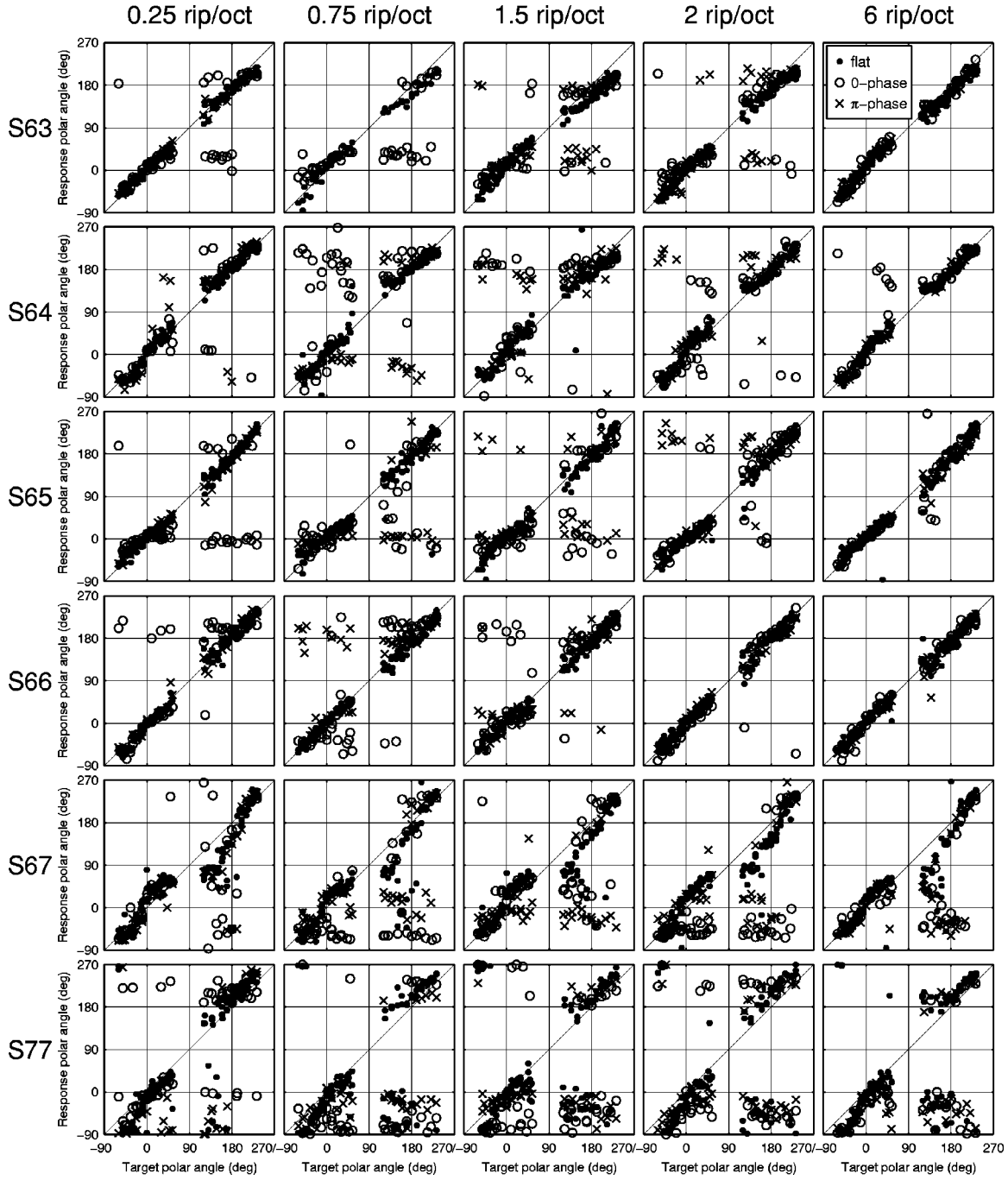


FIG. 4. Response polar angles for selected conditions of Experiment I. Each column of panels contains data for one ripple density. Each row of panels contains data for one listener. Each panel shows the responses for the 0- (open circles) and π-phase ($\times$s) ripple profiles combined with those for the corresponding flat-spectrum stimuli (small closed circles). Ripple depth was 40 dB in all conditions. Data for the 1-ripple/oct condition are plotted in Fig. 8.

for similar analyses in earlier work (Macpherson and Middlebrooks, 2000). Under this criterion, uniformly distributed, random responses would produce an error rate of 75%, and responses falling consistently at 0-deg polar angle would produce an error rate of 62.5% given the range of target polar angles actually used. Depending upon the spatial distribution of errors, error rates of 0%–100% are possible.

Finally, for each block of trials, localization performance for the flat- and ripple-spectrum stimuli was summarized by the *flat-spectrum polar error rate* and the *ripple-spectrum polar error rate*. These were defined as the percentage of responses to the stimuli of each type that were classed as polar errors. For the data shown in Fig. 5, the flat-spectrum polar error rate was 0% and the ripple-spectrum polar error rate was 34%.

c. Polar error rate results. For each of the six listeners, the polar error rates were computed for targets of each ripple density and phase and for the flat-spectrum targets from the corresponding blocks of trials. The results are plotted in Fig. 6. Each panel represents the data for one listener, with polar

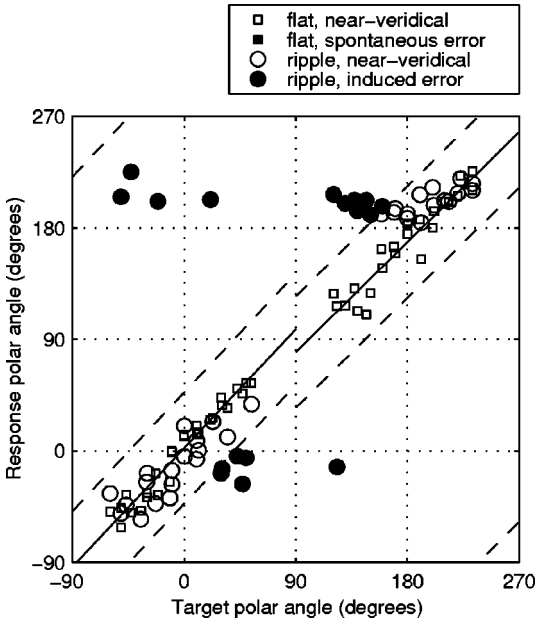


FIG. 5. Sample polar angle response data of listener S66 for 1-ripple/octave, 40 dB-depth, π-phase target spectra. Responses were classified as polar errors (filled symbols) if they fell more than 45 deg away (beyond dashed lines) from a linear function fit (solid lines) to the flat-spectrum responses. Otherwise they were classified as near-veridical (open symbols).

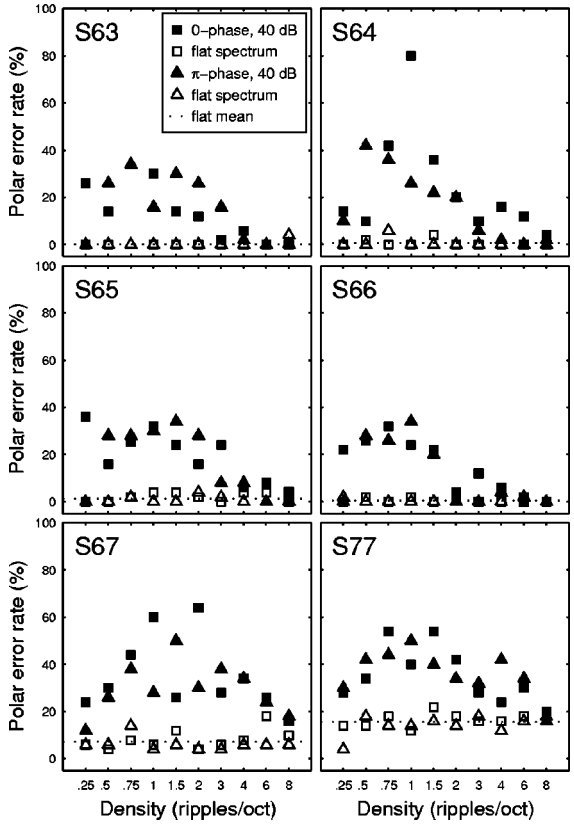


FIG. 6. Results of Experiment I. Ripple-spectrum polar error rates (filled symbols; 0-phase: squares; π-phase: triangles) are plotted as a function of ripple density and phase for each of the six listeners. Flat-spectrum polar error rates (open squares and triangles) are plotted for the flat-spectrum targets presented in the corresponding trial blocks.

error rates (percent) plotted on the ordinate against ripple density (ripples/oct) on the abscissa. Polar error rates for 0- and π-phase ripple stimuli are plotted with closed squares and triangles, respectively. The polar error rates for the flat-spectrum control stimuli presented in association with each type of ripple stimulus are plotted with the corresponding open symbols.

Listeners varied in their flat-spectrum polar error rates. Localization of these targets was almost always near-veridical for four listeners (S63, S64, S65, S66), and hence the mean flat-spectrum polar error rate was low ($<$2%). The remaining two listeners (S67 and S77) showed more frequent flat-spectrum polar errors, and consequently their mean baseline polar error rates were higher ($\sim$7% and 15%, respectively). For those listeners, flat-spectrum errors consisted primarily of rear-to-front reversals. The observed error rates are similar to those described by Makous and Middlebrooks (1990) and Carlile *et al.* (1997). Despite these individual differences, each listener's flat-spectrum polar error rate remained stable across sessions, indicating that localization performance itself did not vary over the course of the experiment and suggesting that differences between error rates for the various ripple stimuli are reliable.

The general effect of adding ripples to the spectral profile was similar across listeners, although the relations between ripple parameters and error rate differed in detail from listener to listener. Listeners with low flat-spectrum error rates (S63, S64, S65, S66) also had the lowest mean ripple-spectrum error rates (mean across all densities for these four, 15%). Those with higher flat-spectrum rates (S67 and S77) had mean ripple error rates about twice as great (33% and 34%, respectively). The highest and most consistent increases in polar-error rate were observed for ripple densities from 0.5 to 2 ripples/octave, in either the 0- or π-phase. In the majority of cases, the ripple-spectrum error rate differed between the phases at a given density within this 0.5–2-ripples/oct range. Computed across listeners and ripple phases, the largest mean ripple-spectrum error rate (37.5%) was produced by the 1-ripple/oct targets. Ripple-spectrum error rates seldom exceeded 50% even for the most disruptive ripple spectra, indicating that the majority of ripple-spectrum targets was localized near-veridically.

At the lowest ripple density, 0.25 ripples/oct, two different patterns of polar error rate increases emerged. For three of the six listeners (S64, S67, S77), both the 0- and π-phase ripples produced modest, but similar, increases in polar error rate. For the other three listeners (S63, S65, S66), the 0-phase ripples resulted in a large increase in error rate, but the π-phase ripples induced no polar errors at all. These two 0.25-ripples/oct profiles are shown in Fig. 1(b), in which it can be seen that the 0-phase profile has a low-pass characteristic, having little energy between 4 and 12 kHz, whereas the π-phase profile lacks energy below 4 kHz.

As ripple density increased above 2 ripples/octave, the ripple-spectrum polar error rate tended to decrease. For three listeners (S63,S65,S66), the ripple-spectrum error rate above 4 ripples/octave was similar to the flat-spectrum error rate. For the other three listeners, the ripple-spectrum error rate declined with increasing ripple density but never reached the

flat-spectrum rate. Two of these listeners, S67 and S77, were those with the highest flat-spectrum error rates.

d. Spatial distribution of polar errors. In addition to the preceding analysis of the polar error rates, we also examined the target angles at which errors were most likely to occur, and the response locations most likely to be produced. Neither flat- nor ripple-spectrum polar errors were distributed uniformly across stimulus location. For all but one of our listeners (S64), errors for both flat- and ripple-spectrum targets were most frequent when targets were presented in the high-rear quadrant. We and others have observed this pattern for flat-spectrum stimuli in previous studies (Carlile *et al.*, 1997; Langendijk *et al.*, 2001; Macpherson and Middlebrooks, 2000; Middlebrooks, 1999b). For S64, however, errors were most common for targets in the high-front quadrant.

In some cases, error and near-veridical responses were *spatially mixed*; both near-veridical and polar error responses were given for targets presented at similar polar angles, as seen in Fig. 5 for targets at polar angles <90 deg. In other cases, error and near-veridical responses were *spatially separated*; many polar errors were induced within particular ranges of target locations while few errors were induced elsewhere. For example, in Fig. 5, almost all targets presented from the high-rear quadrant (90–180 deg) resulted in errors, but none that was presented from the low-rear quadrant (180–270 deg) did.

Although the *likelihood* of polar errors was related to the target location (in that, overall, stimuli from the high-rear quadrant were most likely to generate polar errors), the *locations* to which the listeners pointed when errors occurred (the *error location*) typically did not depend on the target location. Rather, error responses tended to cluster at particular, ripple-dependent polar angles independent of the physical target location. As can be seen in Fig. 4, the scatter in error responses was generally greater than that for near-veridical responses to flat- or ripple-spectrum targets made during the same trial blocks. This suggests that listeners had greater uncertainty about the apparent locations of these stimuli; the auditory images of these ripple stimuli possibly were more diffuse.

e. Distribution of DTF ripple components. Finally, we compared the polar error rate data as a function of ripple density with an estimate of the density distribution of spectral-envelope components (i.e., the *ripple spectrum*) in listeners' DTF profiles. Median-plane DTFs of one listener, S64, are shown in Fig. 7(a). As is visible in this plot, pronounced local features in human DTFs, such as peaks and notches, typically lie at frequencies above 4 kHz, and tend to migrate to higher frequencies as elevation approaches 90 deg both in the front and rear hemispheres. Each of the median-plane DTFs for each listener was resampled on a logarithmic frequency scale, and the transfer-function magnitude spectrum (in dB) was passed through a bank of zero-phase, Gaussian bandpass filters. The center frequencies of these filters varied from 0.25 to 8 ripples/oct in 0.25-ripples/oct increments. All filters had the same bandwidth; the full width at half-maximum was 1 ripple/oct. For each DTF, we defined the bandpass ripple depth as the difference between the highest and lowest values observed over the portion of the filtered spectrum between 1 and 16 kHz. We computed the mean and maximum ripple depth across the set of DTFs in each ripple density channel.

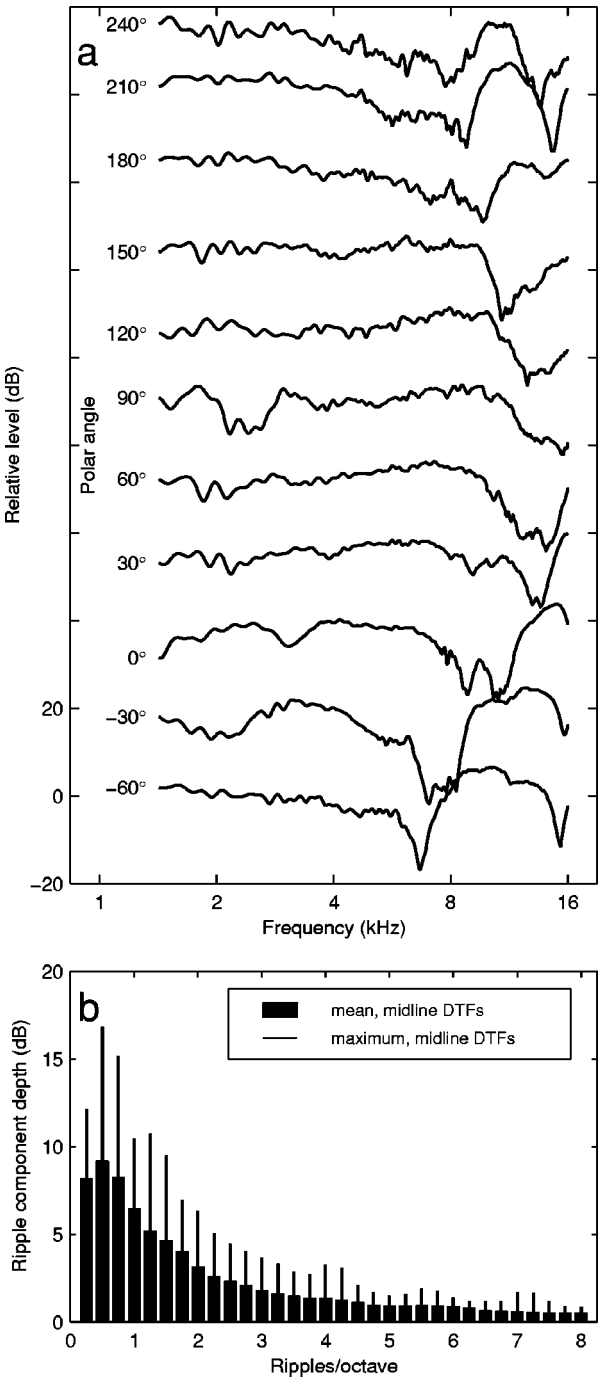


FIG. 7. (a) Median-plane DTFs of listener S64. (b) DTF ripple depth (peak-to-trough ratio in dB) versus ripple frequency after bandpass filtering. Thick bars show the mean depth at each ripple center frequency computed across the set of median-plane DTFs for listener S64. Thin lines show the maximum ripple depth observed across the set.

The resulting DTF ripple spectrum for listener S64 is shown in Fig. 7(b). The ripple spectra for all other listeners were very similar to the one shown except for listener S67, for whom the ripple depth over the range 0.25–1 ripples/oct was lower and approximately constant at ~5 dB. The ripple

spectra for all listeners exhibited a low-pass characteristic, with the highest mean and maximum depths at the lowest ripple densities (0.25–0.75 ripples/oct). There was little ripple energy above 3–4 ripples/oct, and the depths of these high-density components, if presented in isolation, lie well below reported detection thresholds (e.g., Supin *et al.*, 1999).

The observed decrease in the effect on localization performance of high-density ripples agreed with the low-pass character of the DTF ripple spectrum, but the decrease in sensitivity at the lowest ripple densities did not.

Although DTF ripple-component depths were similar among listeners, the details of the component profiles differed, particularly at higher ripple densities. To summarize these differences, we computed, for each median plane DTF measurement location, the Pearson product-moment correlation between the ripple component output of each spectral-envelope bandpass filter for each possible pair of listeners. At 0.25 ripples/oct, the mean correlation across locations and listener pairs was 0.97, at 1 ripple/oct it was 0.72, and above 2 ripples/oct, the mean across frequency channels was ~0.55.

4. Summary

In summary, with a ripple depth of 40 dB, ripple densities of 0.5–2 ripples/octave produced the most pronounced increases in polar error rates for all listeners. Above 2 ripples/oct, the ripple-spectrum polar error rate decreased with increasing ripple density. This was consistent with the lower depths of high-density ripple components in the DTF spectra. At 0.25 ripples/oct, some listeners were able to localize the 0-phase targets accurately, whereas localization by other listeners was degraded by the low-pass character of that stimulus. Vertical-plane localization was disrupted most severely for 1-ripple/oct stimuli. Vertical-plane localization was less sensitive to low-density source-spectrum ripple than predicted from the distribution of ripple components in the DTF spectra.

IV. EXPERIMENT II: EFFECT OF RIPPLE DEPTH

In Experiment II we examined the effect of the depth of modulation of the spectral profile on localization accuracy. In Experiment I, only the rather large 40-dB peak-to-trough ripple depth was used. We hypothesized that smaller ripple amplitudes, particularly amplitudes smaller than those observed in DTFs (Fig. 7), should interfere minimally with the extraction and processing of natural spectral cues for vertical-plane localization. Because the 1-ripple/oct density was found to have the most pronounced effect on polar error rate in Experiment I, the stimulus parameter space in Experiment II was restricted to that density.

A. Stimuli

Target stimuli consisted of ripple-spectrum noise bursts with ripple densities of 1 ripple/oct, phases of 0 or π radians, and peak-to-trough depths of 10, 20, and 30 dB. Blocks of trials were structured in the same manner and used the same target locations as in Experiment 1. The resulting data were combined with the flat-spectrum and 40-dB ripple-depth data previously collected in Experiment I, providing data over a range of ripple depths from 0 to 40 dB.

B. Results

The polar angle components of listeners' responses are plotted versus target polar angle in Fig. 8 for all conditions of Experiment II. Each column of panels contains data for one ripple depth, and each row contains data for one listener. In each panel, responses for the 0- (open circles) and π-phase ($\times$s) ripple profiles are combined with those for the corresponding flat-spectrum stimuli (small closed circles). It can be seen that the number of polar angle errors tended to increase with increasing ripple depth, and that the locations of errors tended to be dependent on ripple phase and independent of target location and ripple depth. At the lowest (10-dB) depth, all listeners except S64 localized the ripple stimuli similarly to the flat-spectrum targets.

The ripple-spectrum error rate for each ripple depth and phase, computed as in Experiment I, are plotted along with the corresponding flat-spectrum error rate in Fig. 9 for all six listeners. For all listeners, the ripple-spectrum error rate was highest at the 40-dB depth and decreased with decreasing ripple depth. For four of the six listeners (S63, S65, S66, and S67), the 10- and 20-dB ripple depths produced similar negligible-to-small increases in the polar error rate. For listeners S64 and S77, the onset of large increases in the polar error rate occurred at a lower ripple depth of 20 dB. As ripple depth was increased, polar errors tended to appear first for targets presented in the high-rear quadrant.

For listeners S65, S66, and S77, the relation between ripple depth and polar error rate was similar for the two ripple phases presented. For S63, S64, and S67, however, the 0-phase ripple profile consistently produced a greater number of polar errors than did the π-phase ripples. This was consistent with the phase-dependent error rates observed for these listeners at 1 ripple/oct and 40-dB depth in Experiment I. For S64 and S67, the polar error rates for the 0 and π ripple phases differed by a factor of 2–3 at the larger ripple depths. In fact, only at the highest ripple depth did the π-phase ripples generate a substantial increase in the polar error rate for these listeners.

Error-response locations did not vary as the depth of the ripples was increased, nor were they dependent on the physical target polar angle. These locations were, however, listener dependent and they varied with the phase of the ripple profile. The one exception to the depth independence of the error-response location was for listener S66 and the 1-ripple/oct, π-phase ripple spectrum, which induced polar errors in the front upper quadrant (i.e., polar angles of 30–60 deg) at a depth of 30 dB, but in the lower-rear quadrant at 40 dB. As in Experiment I, targets presented at similar polar angles could produce both near-veridical and polar error responses in different trials.

In summary, listeners produced near-veridical localization responses for 1-ripple/oct spectra at depths of up to 10 or 20 dB. At higher ripple depths, substantial numbers of polar errors were observed, and their rate of occurrence increased with increasing ripple depth.

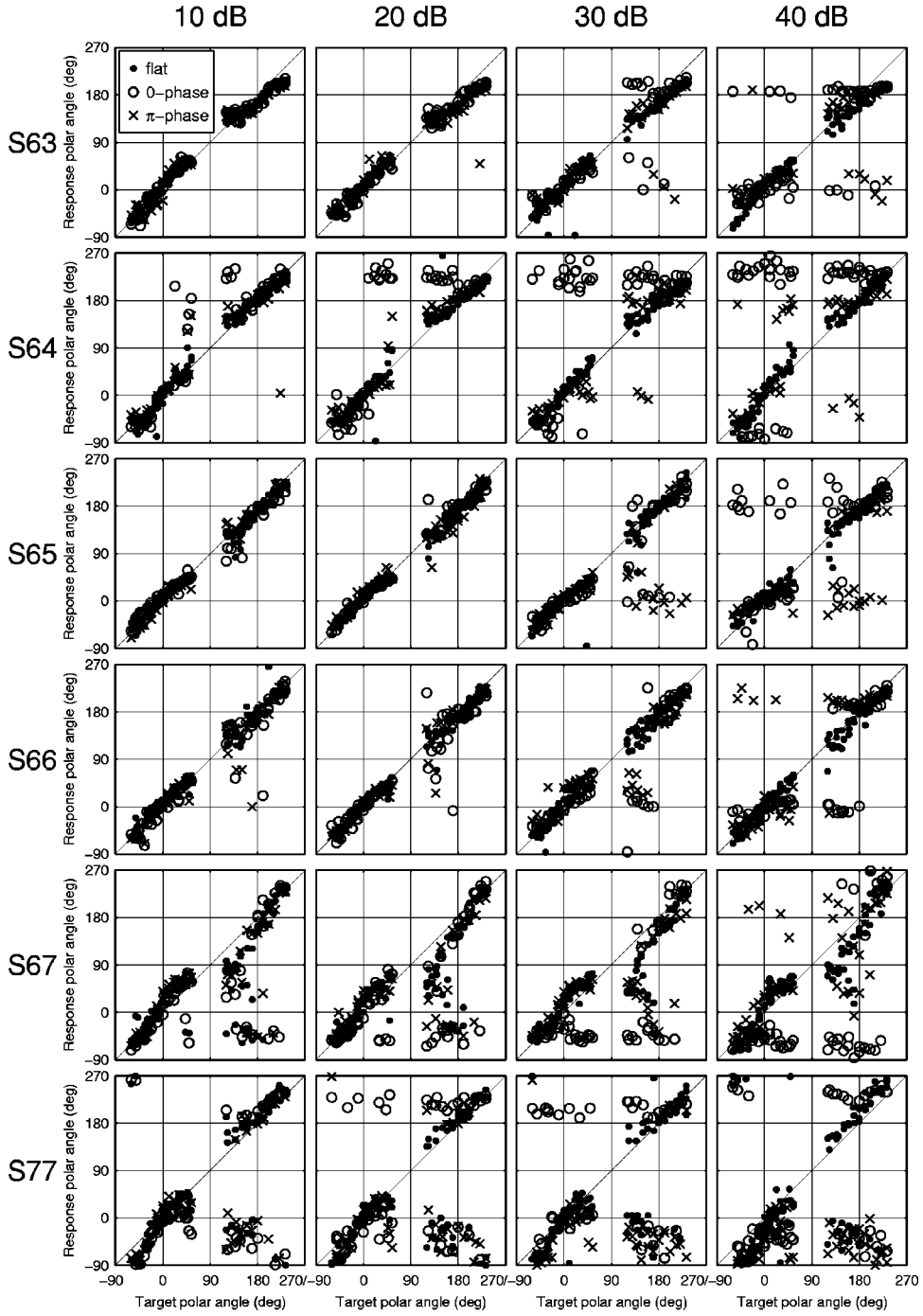


FIG. 8. Response polar angles for all conditions and listeners in Experiment II. Each column of panels contains data for one ripple depth. Each row of panels contains data for one listener. Each panel shows the responses for the 0- (open circles) and π-phase ($\times$s) ripple profiles combined with those for the corresponding flat-spectrum stimuli (small closed circles). Ripple density was 1 ripple/oct in all conditions.

V. EXPERIMENT III: EFFECT OF RIPPLE PHASE

Our hypotheses, that source-spectrum features of *sufficient depth* and of the *appropriate scale* can disrupt vertical-plane localization by obscuring or interfering with DTF cues, are supported by the results of Experiments I and II. In those experiments, we based our conclusions principally on the *rate* of polar-angle localization errors. In Experiment III, we presented more exemplars of the 1-ripple/oct, 40-dB spectra in an attempt to generate sufficient numbers of polar errors for reliable analysis of the relations among the ripple phases, the listeners' DTFs, and the *locations* of polar error responses.

A. Stimuli

Experimental target stimuli consisted of ripple-spectrum noise bursts with a ripple density of 1 ripple/oct, a ripple depth of 40 dB, and phases (the parameter of interest) of $\pm\pi/4$, $\pm\pi/2$, and $\pm 3\pi/4$. Blocks of trials were in all other

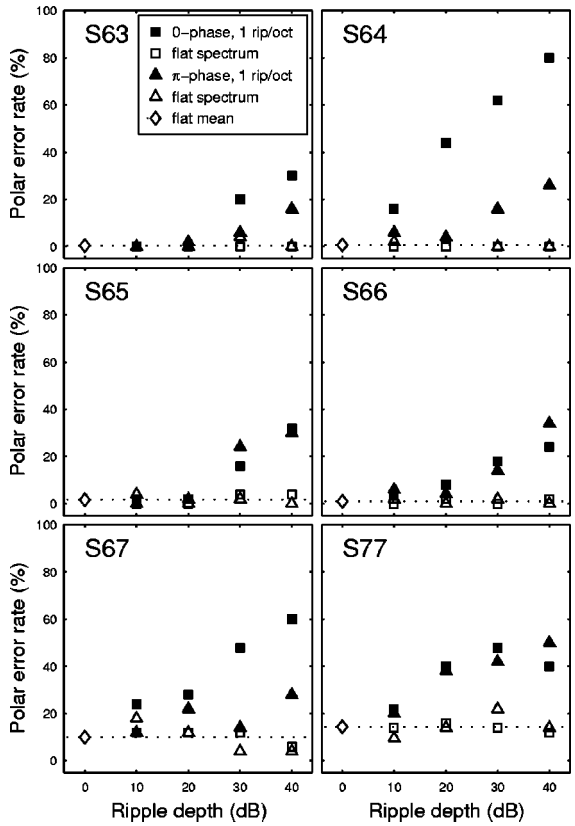


FIG. 9. Results of Experiment II. Ripple-spectrum error rate is plotted as a function of ripple depth and phase for each of the six listeners. All ripple profiles had a density of 1 ripple/oct. Flat-spectrum error rate (open symbols) is plotted for the flat-spectrum targets presented in the corresponding trial blocks. The data for the 40-dB depth were obtained in Experiment I.

respects identical to those used in Experiments I and II. The response data were combined with those for the 0- and π-phase targets presented previously in Experiment I, resulting in a set of eight ripple spectra spanning a 2π-range of phase in increments of $\pi/4$. Data were not collected for listener S65 in the $-3\pi/4$-phase condition because of his limited availability.

B. Results

The polar angle components of listeners' responses are plotted versus target polar angle in Fig. 10 for all conditions of Experiment III. Each column of panels contains data for one ripple phase, and each row contains data for one listener. In each panel, responses for the ripple profiles (open circles) are combined with those for the corresponding flat-spectrum stimuli (small closed circles).

There were marked differences among the listeners in the spatial distribution of polar errors for the ripple-spectrum stimuli, but a similarity was the overall lack of influence of the target polar angle on the locations of errors, as observed in Experiments I and II.

Three of the five listeners (S65, S66, and S77) showed very little modulation of error location as a function of ripple. For S65, the majority of errors occurred for rear-hemisphere targets and error responses typically fell near 0 deg polar angle regardless of ripple phase except for the 0-phase ripples, which produced errors near 180 deg polar angle for front-hemisphere targets. For S66, the dominant error location was also close to 0 deg polar angle. For S77, error locations for most ripple phases were distributed between -90 and 0 deg polar angle. We note that, unlike those of the other listeners, the ripple-spectrum polar errors for S77 were distributed similarly to those for the flat-spectrum targets. The ripple spectra in this case appeared to increase the rate of rear-to-front confusions but not to produce novel error locations.

The data for the remaining two listeners, S64 and S67, were qualitatively different, and showed stronger and more systematic modulation of error location with ripple phase. For S64, the majority of errors fell in the rear hemisphere and moved from locations near 180 deg polar angle for ripple phases near $\pm\pi$ to $\sim$240 deg for 0-phase ripples. A second mode of responses in the front hemisphere followed a similar and symmetrical pattern of elevation (the up/down component of polar angle) change. For listener S67, the majority of error responses fell in the front hemisphere, but error-location polar angle was related similarly to ripple phase as for S64.

For each of the five listeners, flat- and ripple-spectrum error rates were computed for each ripple phase condition as in Experiments I and II. The results are plotted in Fig. 11. For all listeners except S77, who had the highest flat-spectrum error rate, every phase produced a ripple-spectrum error rate higher than the flat-spectrum rate. Polar error rates were modulated by the phase of the ripples differently for each listener. For four listeners (again excluding S77), the differences between the flat-spectrum error rates that were observed in Experiments I and II for the 0- and π-phase ripples are seen to be portions of continuous error-rate-versus-phase functions; this periodicity was predictable because of the periodic dependence of the stimulus spectrum on the ripple phase parameter. For S77, this relation was considerably more irregular. The ripple phase producing the largest error rate varied among listeners. Listeners S64 and S67 had maximum ripple-spectrum error rates for 0-phase ripples. For S66 and S77, the maximum was for the π-phase ripples, and for S65, the maximum was at a phase of $+3\pi/4$. As observed in Experiments I and II, ripple-spectrum targets were localized near-veridically in a majority of trials.

C. Summary

In summary, 1-ripple/oct, 40-dB source spectrum ripples produced listener- and phase-dependent rates of polar errors, but there was no consistent effect of ripple phase on error response locations. For three of the five listeners in Experiment III, error responses fell near specific locations that were not influenced by ripple phase. For the other two listeners, however, polar error response locations were strongly and systematically modulated by the phase of the ripple profile.

VI. DISCUSSION

The results of these experiments support the hypothesis that source-spectrum features within a restricted range of ripple densities can, if sufficiently deep, disrupt vertical-

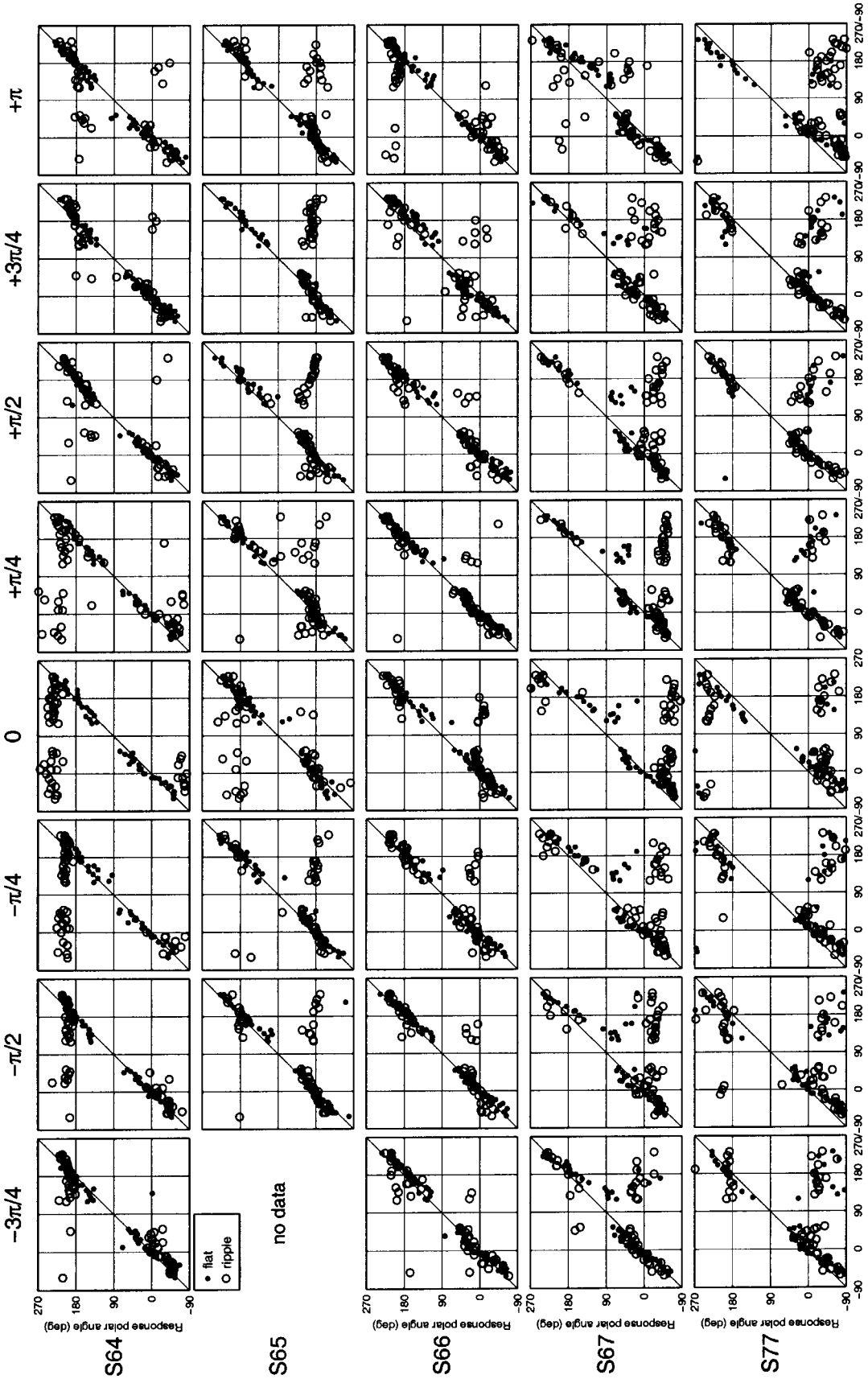


FIG. 10. Response polar angles for all conditions of Experiment III. Each column of panels contains data for one ripple phase. Each row of panels contains data for one listener. Each panel shows the responses for the ripple (open circles) stimuli combined with those for the corresponding flat-spectrum stimuli (small closed circles). In all conditions, ripple density was 1 ripple/oct and ripple depth was 40 dB.

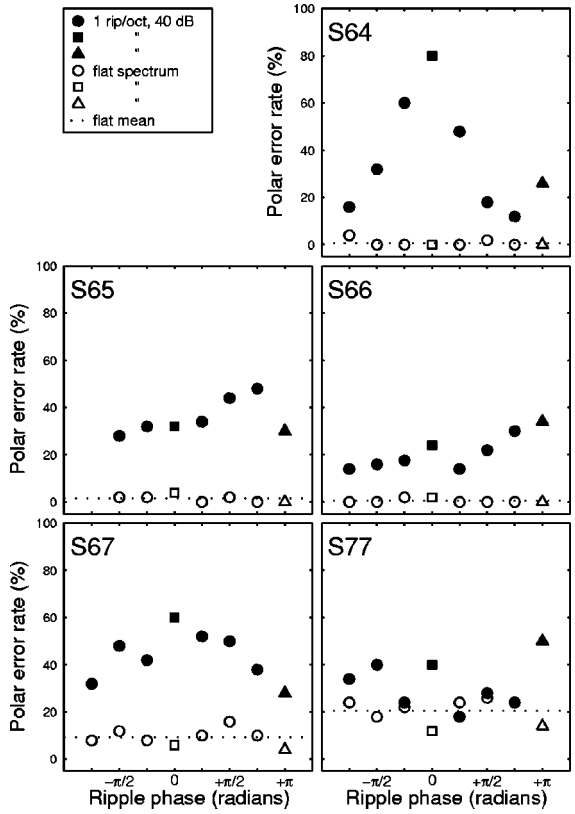


FIG. 11. Ripple-spectrum error rate (filled symbols) is plotted as a function of ripple phase for each of the five listeners in Experiment III. All ripple profiles had a density of 1 ripple/oct and a depth of 40 dB. Flat-spectrum error rate (open symbols) is plotted for the flat-spectrum targets presented in the corresponding trial blocks. Data for 0 (squares) and π (triangles) ripple phases were obtained in Experiment I.

plane localization. Nevertheless, listeners can in many cases extract effective localization cues from the proximal spectra even for source spectra with substantial across-frequency variation within this ripple-density range. The results of the ripple-phase manipulation of Experiment III provided only limited support for the hypothesis that polar errors occur when listeners confuse source-spectrum features with those of DTFs. Some listeners appeared to produce errors near single default locations that were independent of the phase of the ripple spectrum, whereas others made errors at locations that *were* ripple-phase dependent.

A. Ripple density

Rippled spectra with a depth of 40 dB consistently increased the rate of errors in vertical-plane localization only for ripple densities between 0.5 and 2 ripples/octave. Across listeners, vertical plane localization was most prone to ripple-induced errors at the 1-ripple/oct density.

At densities above 2–3 ripples/oct, the rate of polar errors decreased with increasing density. This spectral-envelope frequency cutoff is similar to that found by Kulkarni and Colburn (1998), who asked listeners to discriminate the apparent locations of accurately synthesized virtual auditory space targets from those of targets synthesized with smoothed DTF spectra. They found that listeners could not perform this task until all spectral details with (linear frequency scale) ripple densities greater than ~0.33 ripple/kHz were removed from the smoothed DTFs. At a center frequency of 8 kHz, this corresponds to a ripple-density cutoff of ~2 ripples/oct. The 2-ripples/oct upper cutoff also agrees well with the finding of Langendijk and Bronkhorst (2002) that vertical-plane localization was not disrupted when portions of DTF spectra smaller than 1/2 octave were flattened before filtering target sounds to produce virtual auditory space stimuli.

For three of our six listeners, the ripple-spectrum error rate returned almost to the flat-spectrum error rate at the higher densities, but for the other three, the error rate declined with increasing density but did not reach the flat-spectrum rate. Two of the listeners in the latter group were those with the substantially higher flat-spectrum error rates. We suggest two possible explanations for their elevated ripple-spectrum error rates for high-density ripple stimuli. First, their poor flat-spectrum localization performance might have been caused by over-reliance on nonrobust or irrelevant, high-ripple-frequency information in their DTF spectra. Second, the unusual timbre of the ripple targets might have distracted these listeners, causing them to be less confident about the front–rear location of the rippled targets.

Sensitivity to source-spectrum ripples declined at densities <0.75–1 ripples/oct. This roll-off appeared to be both listener- and ripple-phase-dependent. The low-pass-like, 0-phase, 0.25-ripples/oct profiles induced frequent polar-angle errors for three listeners. For these listeners the results might be explained by the lack of energy in the frequency region above 4 kHz, which is believed to support the primary spectral cues for vertical and front/back localization (e.g., Asano *et al.*, 1990; Gardner, 1973; Hebrank and Wright, 1974). Alternatively, their responses might have been influenced by the high-frequency peak located just below 16 kHz in this spectral profile. The other three listeners must have been able to achieve more accurate performance by obtaining reliable spectral cue information despite the lack of high-frequency energy. For example, they might have been more sensitive than the other listeners to important cues restricted to the attenuated frequency region. For the high-pass-like, π-phase profile, errors were less frequent than for the 0 phase, and three listeners made no errors for these stimuli. In this case, the stimuli had the necessary high-frequency energy, but in order to achieve accurate localization, listeners must also have been able to discount the low-ripple-frequency variation in the target spectra.

Accurate localization of target spectra having low spectral-envelope frequencies has been demonstrated in previous localization experiments (Macpherson, 1998). This robustness could rely on a mechanism which directly filters out gradual variations in spectral slope (Zakarauskas and Cynader, 1993) or on a more general process involving cues extracted only from putative higher ripple-frequency channels (Eddins and Harwell, 2002). If the important spectral features are discrete ones such as notches or sharp spectral gradients, their contribution to the variance in the resulting representation of the spectral profile would likely be increased by either of these processes; suboctave-scale notches and peaks are composed of higher-density ripple components. Langendijk and Bronkhorst (2002) did find that a

template-matching model based on the approach of Zakarauskas and Cynader (1993) did not best predict the responses of listeners to virtual space stimuli synthesized with flattened-spectrum DTFs, but those stimuli did not include deep low-density ripple components.

Lack of sensitivity at low ripple densities is possibly peculiar to the localization process. The spectral resolving power of the auditory system has been measured psychophysically in a number of studies using ripple-detection (Bernstein and Green, 1987; Bero and Eddins, 2001; Green *et al.*, 1987), ripple-phase discrimination (Supin *et al.*, 1994, 1998, 1999), and tone-in-rippled-noise masking tasks (Houtgast, 1977). A common finding is that ripple detection thresholds are lowest and relatively constant for spectral-envelope frequencies below 3–4 ripples/oct. Above 3 ripples/oct, detection and phase-discrimination thresholds increase rapidly with ripple density, but a similar roll-off is not observed at densities <0.5 ripples/oct. These results are consistent with the high-frequency roll-off of ripple sensitivity in our localization task, but not with the low-density roll-off. This suggests a localization-specific specialization of spectral profile processing that discounts DTF spectral features at coarser scales of detail in order to minimize the influence of low-density source-spectrum features.

Because spectrally restricted features such as sharp notches persist in a profile when low-density components are removed, the suggestion that low-ripple-density information is ignored is independent of the issue of whether spectral cues are processed via the recognition of broadband spectral templates or of discrete spectral features. One possibility is that localization involves discrete-feature detectors that are intrinsically insensitive to low-density ripple components. A second possibility is that a first stage of processing removes low-density variation and a second stage operates on this spectral representation either by template matching or by detection of discrete features.

Several recent studies (Eddins and Harwell, 2002; Saoji and Eddins, 2002) support the notion that the auditory system possesses spectral-envelope frequency channels analogous to the spatial frequency channels found in the visual system. The bandpass character of the sensitivity of localization to source-spectrum ripples is consistent with this proposal given the assumption that spectral-cue extraction is performed primarily in spectral-envelope frequency channels centered near 1 ripple/oct. The best evidence is again the lack of sensitivity at very low ripple densities, as lack of sensitivity at higher densities is likely simply a result of peripheral filtering. Saoji and Eddins (2002) describe a ''spectral modulation detection interference'' phenomenon similar to the interference we have assumed takes place between the source-spectrum ripple and spectral features of the DTFs. The fact that we observe most interference with localization at 1 ripple/oct suggests that this is the key scale at which the auditory system analyzes spectral cues for vertical plane localization.

The high-density (low-pass) cutoff of localization sensitivity to source-spectrum ripples is consistent with the ripple analysis of DTF spectral profiles (Fig. 7) and with the decline in the detectability of high-density ripples, but the low-density cutoff is not. As shown in Fig. 7, DTF ripple spectra show a low-pass characteristic, with the strongest component measured at 0.25 ripples/oct and little energy above 3–4 ripples/octave, but DTF ripple spectra and ripple detectability do not show a low-ripple-frequency roll-off that would correspond with our psychophysical result. This demonstrates that our results, at least for ripple densities <0.5 ripples/oct, are not a simple consequence of the acoustical properties of the DTFs or of spectral smoothing in the cochlea. The most disruptive density (1 ripple/oct), however, roughly corresponds to the frequency separation of deep, high-frequency notches and their flanking peaks that we observed in our listeners' DTFs. By inspection of DTF plots similar to Fig. 7 for each listener, we determined that such features were typically separated by 0.7–1.5 octaves.

The observed bandpass-like psychophysical sensitivity of polar-angle localization to ripple density closely matches the distribution of ripple-tuned cells found in a study of the auditory cortex of the cat (Shamma *et al.*, 1995). In that study, 74% of the sampled units were tuned to ripple frequencies between 0.25 and 2 ripples/oct. Ripple densities >3 ripples/oct were not tested in that study. Whether human auditory cortex displays a similar distribution of ripple tuning is unknown. The most prominent spectral features of the DTFs of cats lie at frequencies higher than those of humans, but the separation of peaks and notches (and hence the spectral scale) of these features in the two species is not dissimilar (Huang and May, 1996; Musicant *et al.*, 1990; Rice *et al.*, 1992; Xu and Middlebrooks, 2000).

B. Ripple depth

We found that the depth of 1-ripple/oct target spectrum ripples had to be at least 20 to 30 dB (depending upon the ripple phase) to substantially increase the rate of polar errors. As shown in Fig. 7(a), median-plane DTF spectra have peak-to-trough ratios that can approach 20–25 dB at some locations. When only ripple components near 1 ripple/oct are considered, the maximum peak-to-trough ratio is $\sim$10 dB [Fig. 7(b)]. Thus, it seems the amplitude of the ripple profile must approach or exceed the extrema in the DTF spectra before substantial interference with localization takes place. We note that even at the lowest depth used (10 dB), the ripple profiles in Experiment II were modulated well above the detection threshold at 1 ripple/oct, which, when expressed as peak-to-trough ratio in dB, has been found by several authors to be 1–2 dB (Bernstein and Green, 1987; Green *et al.*, 1987; Supin *et al.*, 1999). Even at a 40-dB ripple depth, near-veridical localization responses occurred in a majority of trials. The extraction of spectral cues from nonflat wideband source spectra appears to be a surprisingly robust process, and one which has yet to be successfully modeled, as previously noted (Macpherson, 1997, 1998).

Earlier studies in which the effects of various depths of source-spectrum variation are reported have been ones in which target spectrum levels were roved independently in 1/3-octave or critical bands. Considering pairs of adjacent 1/3-octave bands as single cycles at a scale of 1.5 ripples/oct, it is clear that such scrambled spectra contain substantial ripple components within the critical region (0.5–2 ripples/

oct) that we identified in Experiment I. Kulkarni and Colburn (1998) reported that roving band level within a 10-dB range had no adverse effect on median-plane localization accuracy. Wightman and Kistler (1997) observed increases in front–back and elevation errors above the flat-spectrum rate with a 20-dB rove and further increases when the rove was increased to 40 dB. These results are in general agreement with our own, although it is not possible to exactly equate the ranges of random spectral scrambling with the depths of our sinusoidal ripple profiles.

The general lack of depth- and target-location-dependence of the locations of polar errors suggests that ripple-spectrum errors did not arise from fortuitous combinations of the spectral features of the target ripple spectra and the target-location DTFs, but from the gradual obscuring (and possibly replacement) of the natural DTF cues as the ripple depth was increased. This interpretation is also supported by the observation that, for most listeners, ripple-spectrum polar errors appeared at the lowest ripple depths in the high-rear quadrant, where DTF features are least pronounced. Errors occurred in other quadrants at higher ripple depths.

C. Ripple phase

In Experiment III, we obtained localization judgments for a variety of phases of the 1-ripple/oct ripple spectrum, which was found to be the most disruptive to accurate localization in Experiment I. The rate of ripple-spectrum polar errors was ripple-phase- and listener dependent. Because individuals' DTFs differ primarily in the frequency loci at which characteristic spectral features lie (Middlebrooks, 1999a), there is no reason to expect *a priori* that a single ripple phase would induce more polar errors than another for all listeners. Instead, particular acoustical interactions of source ripples and DTF spectra (for example, the filling-in of a DTF notch by a ripple peak or the attenuation of a DTF peak by a ripple trough) should be expected at different ripple phases for different individuals. Because DTF ripple components vary more among individuals at 1 ripple/oct then at 0.25 ripples/oct (Sec. III B 3 e), it is not unexpected that the effects of phase on error rates in Experiment III were more idiosyncratic than those observed in Experiment I at 0.25 ripples/oct.

The exact shapes of the error-rate-versus-phase functions (Fig. 11) are not as instructive as are the facts that there *was* variation with ripple phase and that the functions were listener dependent. The modulation of error rate by ripple phase indicates that the details of the source-spectrum profile were important contributors to the number and, for some listeners, the location of polar errors. It is not simply the case that a ripple spectrum of a particular density and depth completely obscured the target-location DTF, causing the listener to make haphazard responses. Instead, errors tended to cluster at particular ripple-dependent polar angles or at consistent ripple-independent locations. The latter outcome is perhaps evidence of a default response produced by some listeners in the face of ambiguity in the available spectral cues. It is not clear, however, why this should have occurred on some trials when on other trials (typically in the majority), near-veridical responses were made for the same source spectrum presented at nearby target locations.

A listener's near-veridical responses to the various ripple stimuli presumably involve the extraction of adequate spectral cues from the proximal spectrum and the generation of a response mediated by the auditory system's representation of the listener's own DTFs (Langendijk and Bronkhorst, 2002; Middlebrooks, 1999b). In Experiment III, the fact that locations of polar errors for listeners S65, S66, and S77 were mostly phase independent suggests that source-spectrum features were not mistaken for DTF features for these listeners. Rather, it seems that the ripple spectra obscured the DTF information and elicited default responses (S65 and S66) or increased rear-to-front confusions (S77).

For S64 and S67, however, the results are suggestive of the interpretation of the proximal spectrum as if it were a natural DTF. As ripple phase increases, the peak–trough pattern in the ripple profile shifts downwards in frequency. For these listeners, the systematic movement of error responses below the horizontal plane as ripple phase increases from $-3\pi/4$ to 0 might be related to the downward shift in the frequency loci of high-frequency DTF notches and peaks as location changes from above to below the horizontal plane. This can be seen in both the front and rear hemispheres in Fig. 7. It is difficult, however, to show an exact correspondence between features of the proximal spectra and those of the DTFs corresponding to the locations at which error responses are made, and the increase in response elevation as ripple phase increased beyond 0 cannot be explained in this way.

VII. CONCLUSIONS

In summary, at a peak-to-trough ripple depth of 40 dB, polar errors of localization were most consistently induced for ripple densities of 0.5–2 ripples/oct. Across listeners, localization performance was most adversely affected by 1-ripple/oct target spectra. The lack of sensitivity to ripple densities above 2–3 ripples/oct can be accounted for by a combination of limited spectral resolution in the cochlear filter bank and the relative lack of fine-scale detail in DTFs themselves. Below 0.5 ripples/oct, it appears that listeners have some mechanism to discount coarse source-spectrum features. This mechanism is perhaps related to analysis of source- and DTF spectra in a restricted range of spectral-envelope frequency bands.

Ripple depths <20 dB at a density of 1 ripple/oct rarely caused increased rates of polar angle localization errors. This suggests that errors occur only when the depth of the source-spectrum features exceeds that of those in the DTF. Even at larger ripple depths, however, localization was still accurate in a majority of trials. Thus, vertical-plane localization displays a remarkable robustness to across-frequency variance in source spectra that has yet to be explained.

ACKNOWLEDGMENTS

The authors are grateful to Leslie Bernstein, Erno Langendijk, and an anonymous reviewer for their constructive

comments on an earlier version of this paper. Christopher Stecker, Ian Harrington, Matt Charous, and Brian Mickey also provided helpful suggestions. Abhijit Kulkarni provided important information about stimulus manipulations in a previous study. Zekiye Onsan provided invaluable technical assistance. Work supported by NIH Grants R01DC00420 and T32DC00011.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). ''Role of spectral cues in median plane localization,'' J. Acoust. Soc. Am. **88**, 159–168.

Bernstein, L. R., and Green, D. M. (**1987**). ''The profile-analysis bandwidth,'' J. Acoust. Soc. Am. **81**, 1888–1895.

Bero, E. M., and Eddins, D. A. (**2001**). ''Spectral envelope perception: Auditory transfer functions for sinusoidal spectral ripple,'' in Abstracts of the 24th Annual Midwinter Meeting (Assoc. Res. Oto.), p. 89(A).

Blauert, J. (**1969**/**70**). ''Sound localization in the median plane,'' Acustica **22**, 205–213.

Burger, J. F. (**1958**). ''Front–back discrimination of the hearing system,'' Acustica **8**, 301–302.

Campbell, F. W., and Robson, J. G. (**1968**). ''Application of Fourier analysis to the visibility of gratings,'' J. Physiol. (London) **197**, 551–566.

Carlile, S., Leong, P., and Hyams, S. (**1997**). ''The nature and distribution of errors in sound localization by human listeners,'' Hear. Res. **114**, 179–196.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (**1982**). ''Spatial frequency selectivity of cells in macaque visual cortex,'' Vision Res. **22**, 545–559.

Eddins, D. A., and Bero, E. M. (**2001**). ''Just noticeable differences in spectral envelope frequency,'' in Abstracts of the 24th Annual Midwinter Meeting (Assoc. Res. Oto.), p. 89(A).

Eddins, D. A., and Harwell, R. M. (**2002**). ''Spatial frequency channels in audition?'' in Abstracts of the 25th Annual Midwinter Meeting (Assoc. Res. Oto.), p. 197(A).

Gardner, M. B. (**1973**). ''Some monaural and binaural facets of median plane localization,'' J. Acoust. Soc. Am. **54**, 1489–1495.

Gilkey, R. H., Good, M., Ericson, M. A., Brinkman, J., and Stewart, J. M. (**1995**). ''A pointing technique for rapidly collecting localization responses in auditory research,'' Behav. Res. Methods Instrum. Comput. **27**, 1–11.

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.

Hammershøi, D., and Møller, H. (**1996**). ''Sound transmission to and within the human ear canal,'' J. Acoust. Soc. Am. **100**, 408–427.

Hebrank, J. H., and Wright, D. (**1974**). ''Spectral cues used in the localization of sound sources on the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834.

Hofman, P. M., and Van Opstal, A. J. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.

Houtgast, T. (**1977**). ''Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker,'' J. Acoust. Soc. Am. **62**, 409–415.

Huang, A. Y., and May, B. J. (**1996**). ''Sound orientation behavior in cats. II. Mid-frequency spectral cues for sound localization,'' J. Acoust. Soc. Am. **100**, 1070–1080.

Kulkarni, A., and Colburn, H. S. (**1998**). ''Role of spectral detail in sound-source localization,'' Nature (London) **396**, 747–749.

Langendijk, E. H. A., and Bronkhorst, A. W. (**2002**). ''Contribution of spectral cues to human sound localization,'' J. Acoust. Soc. Am. **112**, 1583–1596.

Langendijk, E. H. A., Kistler, D. J., and Wightman, F. L. (**2001**). ''Sound localization in the presence of one or more distracters,'' J. Acoust. Soc. Am. **109**, 2123–2134.

Macpherson, E. A. (**1997**). ''A comparison of spectral correlation and local feature-matching models of pinna cue processing,'' J. Acoust. Soc. Am. **101**, 3104(A).

Macpherson, E. A. (**1998**). ''Spectral cue processing in the auditory localization of sounds with wideband non-flat spectra,'' Ph.D. dissertation, University of Wisconsin-Madison.

Macpherson, E. A., and Middlebrooks, J. C. (**2000**). ''Localization of brief sounds: Effects of level and background noise,'' J. Acoust. Soc. Am. **108**, 1834–1849.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). ''Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,'' J. Acoust. Soc. Am. **111**, 2219–2236.

Makous, J. C., and Middlebrooks, J. C. (**1990**). ''Two-dimensional sound localization by human listeners,'' J. Acoust. Soc. Am. **87**, 2188–2200.

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C. (**1999a**). ''Individual differences in external-ear transfer functions reduced by scaling in frequency,'' J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C. (**1999b**). ''Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency,'' J. Acoust. Soc. Am. **106**, 1493–1510.

Middlebrooks, J. C., and Green, D. M. (**1990**). ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**, 2149–2162.

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' in *Annual Review of Psychology* (Annual Reviews, Palo Alto), pp. 135–159.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). ''Directional sensitivity of sound-pressure levels in the human ear canal,'' J. Acoust. Soc. Am. **86**, 89–108.

Musicant, A. D., Chan, J. C. K., and Hind, J. E. (**1990**). ''Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons,'' J. Acoust. Soc. Am. **87**, 757–781.

Pierce, A. H. (**1901**). *Studies in Auditory and Visual Space Perception* (Longmans Green, New York).

Rakerd, B., Hartmann, W. M., and McCaskey, T. L. (**1999**). ''Identification and localization of sound sources in the median sagittal plane,'' J. Acoust. Soc. Am. **106**, 2812–2820.

Rice, J. J., May, B. J., Spirou, G. A., and Young, E. D. (**1992**). ''Pinna-based spectral cues for sound localization in cat,'' Hear. Res. **58**, 132–152.

Saoji, A. A., and Eddins, D. A. (**2002**). ''Monaural and binaural spectral modulation detection interference,'' in Abstracts of the 25th Annual Midwinter Meeting (Assoc. Res. Oto.), p. 21(A).

Shamma, S. A., Versnel, H., and Kowalski, N. (**1995**). ''Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra,'' Aud. Neurosci. **1**, 233–254.

Shinn-Cunningham, B., and Kopčo, N. (**2002**). ''Effects of reverberation on spatial auditory performance and spatial auditory cues,'' J. Acoust. Soc. Am. **111**, 2440(A).

Strutt, J. W. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.

Supin, A. Y., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1994**). ''Frequency resolving power measured by rippled noise,'' Hear. Res. **78**, 31–40.

Supin, A. Y., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1998**). ''Ripple density resolution for various rippled-noise patterns,'' J. Acoust. Soc. Am. **103**, 2042–2050.

Supin, A. Y., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1999**). ''Ripple depth and density resolution of rippled noise,'' J. Acoust. Soc. Am. **106**, 2800–2804.

Versnel, H., and Shamma, S. A. (**1998**). ''Spectral-ripple representation of steady-state vowels in primary auditory cortex,'' J. Acoust. Soc. Am. **103**, 2502–2514.

Wightman, F. L., and Kistler, D. J. (**1989**). ''Headphone simulation of free-field listening. I. Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1997**). ''Monaural localization revisited,'' J. Acoust. Soc. Am. **101**, 1050–1063.

Xu, L., and Middlebrooks, J. C. (**2000**). ''Individual differences in external-ear transfer functions of cats,'' J. Acoust. Soc. Am. **107**, 1451–1459.

Zahorik, P., Wightman, F. L., and Kistler, D. J. (**1995**). ''On the discriminability of virtual and real sound sources,'' in *Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE, New York).

Zakarauskas, P., and Cynader, M. S. (**1993**). ''A computational theory of spectral cue localization,'' J. Acoust. Soc. Am. **94**, 1323–1331.

Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). ''Characterization of external ear impulse responses using Golay codes,'' J. Acoust. Soc. Am. **92**, 1169–1171.

Zwicker, E., and Terhardt, E. (**1980**). ''Analytical expressions for critical-band rate and critical bandwidth as a function of frequency,'' J. Acoust. Soc. Am. **68**, 1523–1525.

Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers

Michael K. Qin[a] and Andrew J. Oxenham[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

(Received 8 November 2002; revised 30 March 2003; accepted 14 April 2003)

This study investigated the effects of simulated cochlear-implant processing on speech reception in a variety of complex masking situations. Speech recognition was measured as a function of target-to-masker ratio, processing condition (4, 8, 24 channels, and unprocessed) and masker type (speech-shaped noise, amplitude-modulated speech-shaped noise, single male talker, and single female talker). The results showed that simulated implant processing was more detrimental to speech reception in fluctuating interference than in steady-state noise. Performance in the 24-channel processing condition was substantially poorer than in the unprocessed condition, despite the comparable representation of the spectral envelope. The detrimental effects of simulated implant processing in fluctuating maskers, even with large numbers of channels, may be due to the reduction in the pitch cues used in sound source segregation, which are normally carried by the peripherally resolved low-frequency harmonics and the temporal fine structure. The results suggest that using steady-state noise to test speech intelligibility may underestimate the difficulties experienced by cochlear-implant users in fluctuating acoustic backgrounds.
[DOI: 10.1121/1.1579009]

PACS numbers: 43.66.Ts, 43.71.Ky, 43.66.Dc [PA]

I. INTRODUCTION

Speech has been shown to be a very robust medium for communicating information (Fletcher and Galt, 1950; Miller and Licklider, 1950; Remez *et al.*, 1994; Stevens, 1998). Although the precise mechanisms underlying the apparent resilience to interference and distortion are still not well understood, the ability of speech to convey information under adverse conditions is generally attributed to the layers of acoustic, phonetic, and linguistic redundancies. Shannon *et al.* (1995), using a noise-excited vocoder,[1] provided a dramatic demonstration of these redundancies at work. They found that despite a severe reduction in spectral cues and the elimination of temporal fine-structure information, sentences presented in the absence of interfering sounds could be recognized with as few as four frequency bands. Subsequent studies have shown that while more frequency bands are needed for speech reception in steady-state noise, good sentence recognition is still possible at relatively low signal-to-noise ratios (e.g., Dorman *et al.*, 1998).

The processing schemes used in these studies are designed to simulate the effects of cochlear-implant stimulation (Wilson *et al.*, 1991). They can therefore be used to provide insights into the relative efficacy of different processing algorithms without using valuable implantee testing time (Blamey *et al.*, 1984). Indeed, at least for low numbers of frequency bands, acoustic simulations of cochlear-implant processing using normal-hearing listeners have yielded results that are reasonably comparable to those of actual implant patients (Friesen *et al.*, 2001; Carlyon *et al.*, 2002). Another use for such schemes is to probe the acoustic features necessary for speech reception in normal-hearing listeners. A number of studies indicate that important information is carried in the envelopes of the stimulus after filtering into frequency sub-bands (Houtgast *et al.*, 1980; Drullman, 1995; Smith *et al.*, 2002). From the results obtained so far, it may be concluded that speech reception requires minimal frequency selectivity and no temporal fine-structure information. This conclusion seems at odds with the experiences of many hearing-impaired listeners.

While hearing-impaired listeners often perform well in quiet conditions (when audibility is corrected for with amplification), many experience great difficulty in noisy conditions. The difference in performance between normal-hearing and hearing-impaired listeners is especially pronounced in temporally fluctuating maskers and maskers with spectral notches (Festen and Plomp, 1990; Gustafsson and Arlinger, 1994; Peters *et al.*, 1998). In particular, while normal-hearing listeners show large improvements in speech reception when spectral and/or temporal fluctuations are introduced into a masker, hearing-impaired listeners often show much less benefit (Festen and Plomp, 1990; Peters *et al.*, 1998). It is thought that normal-hearing listeners are able to make use of the improved local target-to-masker ratio in the masker's spectral and temporal dips. In contrast, hearing-impaired listeners, with their poorer frequency selectivity (Patterson *et al.*, 1982; Glasberg and Moore, 1986) and poorer effective temporal resolution (Glasberg and Moore, 1992; Oxenham and Moore, 1997), may be less able to benefit from the improved local target-to-masker ratio found in the spectral and temporal dips of the masker.

[a]Electronic mail: qin@mit.edu
[b]Electronic mail: oxenham@mit.edu

 0001-4966/2003/114(1)/446/9/$19.00

In the case of cochlear implants and implant simulations, the finding that better frequency resolution (i.e., a greater number of frequency bands) is required for speech reception in noise than in quiet (Dorman *et al.*, 1998; Fu *et al.*, 1998) parallels the finding that spectral smearing is more detrimental to speech reception in noise than in quiet (ter Keurs *et al.*, 1992; Baer and Moore, 1993). It is also consistent with the hypothesized effect of poorer frequency selectivity in hearing-impaired listeners. The perceptual effect of eliminating the temporal fine structure in cochlear-implant simulations is less clear. Pitch perception and the ability to discriminate different fundamental frequencies ($F0$s), is thought to rely primarily on fine-structure information, in particular the information carried in peripherally resolved, lower-order harmonics (e.g., Plomp, 1967; Houtsma and Smurzynski, 1990; Smith *et al.*, 2002). While the envelopes of implant-processed stimuli carry some periodicity information, the pitch salience associated with such envelope periodicity is rather weak (Burns and Viemeister, 1976; 1981; Shackleton and Carlyon, 1994).

Fundamental frequency information has long been thought to play an important role in perceptually segregating simultaneous and nonsimultaneous sources (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990; 1994; Bird and Darwin, 1998; Vliegen and Oxenham, 1999; see Darwin and Carlyon, 1995 for a review). A reduction in $F0$ cues produced by cochlear-implant processing may lead to greater difficulties in segregating different sources. If the perception of implant-processed speech is based on envelope fluctuations, as suggested above, then listeners must accurately distinguish the envelope fluctuations of the target from those of the masker. Similarly, a listener can only take advantage of spectral and temporal dips in the masker if the listener can accurately identify the presence of the dips.

The aim of the present study was to investigate the effects of fluctuating maskers on the reception of simulated implant-processed speech. We hypothesized that the reduction in $F0$ cues produced by the implant simulations would particularly affect conditions where the ability to discriminate the target from the masker is thought to play an important role in determining speech reception thresholds (e.g., speech in the presence of competing talkers or fluctuating backgrounds). Speech reception was measured in normal-hearing listeners as a function of target-to-masker ratio, processing condition (4, 8, or 24 channels, or unprocessed) and masker type (steady-state speech-shaped noise, speech-shaped noise modulated with a speech envelope, single male talker, and single female talker).

II. METHODS

A. Subjects

Thirty-two normal-hearing listeners (15 females) with audiometric thresholds of less than 20 dB HL at octave frequencies between 125 and 8000 Hz, participated in this study. Their ages ranged from 18 to 46 (median age 22). They were all native speakers of American English.

B. Stimuli

All stimuli in this study were composed of a target sentence presented in the presence of a masker. The stimulus tokens were processed prior to each experiment. The targets and maskers were combined at the desired target-to-masker ratios (TMRs) prior to any processing. TMRs were computed based on the token-length root-mean-square (rms) amplitudes of the signals. Maskers were gated on and off with 250-ms raised-cosine ramps 250 ms prior to and 250 ms after the end of each target sentence.

The targets were H.I.N.T. sentences (Nilsson *et al.*, 1994) spoken by a male talker. The H.I.N.T sentence corpus consists of 260 phonetically balanced high-context sentences of easy-to-moderate difficulty. Each sentence is composed of four to seven keywords.

Since differences in the $F0$ of voicing are thought to contribute to speaker segregation (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990; 1994; Darwin and Carlyon, 1995; Bird and Darwin, 1998), we chose a male single-talker masker with a mean $F0$ (111.4 Hz) similar to that of the target talker (110.8 Hz) and a female single-talker masker with a mean $F0$ (129.4 Hz) almost 3 semitones higher. The motivation for using different gender single-talker interferers came from the observation that normal-hearing listeners benefit from $F0$ differences between target and interfering talkers (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990; 1994; Bird and Darwin, 1998). Talker $F0$s were estimated using the YIN program provided by de Cheveigné and Kawahara (2002). The male single-talker maskers were excerpts from the audio book *Timeline* (novel by M. Crichton) read by Stephen Lang. The female single-talker maskers were excerpts from the audio book *Violin* (novel by A. Rice) read by Maria Tucci. To avoid long silent intervals in the masking speech, such as sentence-level pauses, both single-talker maskers were automatically preprocessed to remove silent intervals greater than 100 ms. The maskers were then subdivided into nonoverlapping segments to be presented at each trial.

The single-talker maskers and speech-shaped-noise masker were spectrally shaped to match the long-term power spectrum of the H.I.N.T. sentences. The amplitude-modulated speech-shaped noise masker was generated by amplitude modulating the steady-state speech-shaped noise with the broadband speech envelope of the male single-talker masker (lowpass filtered at 50 Hz; first-order Butterworth filter).

For a given listener, the target sentence lists were chosen at random, without replacement, from among the 25 lists of H.I.N.T. sentences. This was done to ensure that no target sentence was presented more than once to any given listener. Data were collected using one list (i.e., ten sentences) for each TMR.

C. Stimulus processing

All stimulus tokens were processed prior to each experiment. The cochlear-implant simulator was implemented using MATLAB (Mathworks, Natick, MA) in the following manner. The stimuli (target plus masker) were first bandpass

TABLE I. Target-to-masker ratios (TMR) used in the study. The values in the table represent the minimum, maximum, and step size of the TMR (in dB). The step sizes are in parentheses.

Processing condition	Masker type	Target-to-masker ratio (dB)
Unprocessed	Male interference	−20 to 5 (5)
	Female interference	−20 to 5 (5)
	Modulated noise	−25 to 0 (5)
	Steady-state noise	−15 to 0 (3)
24 channels	Male interference	−15 to 10 (5)
	Female interference	−15 to 10 (5)
	Modulated noise	−20 to 5 (5)
	Steady-state noise	−10 to 10 (4)
8 channels	Male interference	−5 to 20 (5)
	Female interference	−5 to 20 (5)
	Modulated noise	−10 to 15 (5)
	Steady-state noise	−5 to 20 (5)
4 channels	Male interference	5 to 30 (5)
	Female interference	5 to 30 (5)
	Modulated noise	5 to 30 (5)
	Steady-state noise	5 to 30 (5)

filtered (sixth-order Butterworth filters) into 4, 8, or 24 contiguous frequency bands (or channels) between 80 and 6000 Hz. The entire frequency range was divided equally in terms of the Cam scale[2] (Glasberg and Moore, 1990). The 3-dB channel bandwidths were approximately 6.98 Cams, 3.49 Cams, and 1.16 Cams for the 4-, 8-, and 24-channel conditions, respectively. The envelopes of the signals were extracted by half-wave rectification and lowpass filtering (using a second-order Butterworth filter) at 300 Hz, or half the bandpass filter bandwidth, whichever was lower. The 300-Hz cutoff frequency was chosen to preserve, as far as possible, $F0$ cues in the envelope. The envelopes were then used to modulate narrowband noises, filtered by the same bandpass filters that were used to filter the original stimuli. Finally, the modulated narrowband noises were summed and scaled to have the same level as the original stimuli.

D. Procedure

The 32 listeners were divided into four groups of eight. Each group was tested on only one of the four processing conditions (i.e., 4, 8, 24 channels, or unprocessed). The speech reception of each listener was measured in the presence of all four masker types (single male and female talkers, modulated and steady-state speech-shaped noise), at six TMRs (see Table I). The TMRs for each processing condition and masker type were determined in an earlier pilot study, using two to three listeners. The TMRs for each experimental condition were set to avoid floor and ceiling effects in the psychometric function.

The target and masker were combined at the appropriate TMR, processed, and stored on disk prior to the experiments. The processed stimuli were converted to the analog domain using a soundcard (LynxStudio, LynxOne) at 16-bit resolution with a sampling rate of 22 050 Hz. The stimuli were then passed through a headphone buffer (TDT HB6) and presented diotically at 60 dB SPL via Sennheiser HD580 headphones to the listener seated in a double-walled sound-insulation booth. Listeners typed their responses into a computer via the keyboard.

For practice, the listeners were presented with 20 stimuli, five from each of the four masking conditions. In each practice masking condition, the target sentences were presented at four different TMRs. The target sentences used in the practice session were from the Harvard-Sentence database (IEEE, 1969). The practice sessions were designed to acclimate the listeners to the processed stimuli. Feedback was given during the practice sessions, but not during the experimental sessions.

E. Analysis

Listener responses were scored offline by the experimenter. Each listener's responses for a given TMR, under a given masker condition, were grouped together to produce a percent-correct score. Keywords were used to calculate the percent correct. Obvious misspellings of the correct word were considered correct.

III. RESULTS

A. Fits to the psychometric functions

The percent-correct scores as a function of TMR under a given masker condition for each listener were fitted to a two-parameter sigmoid model (a cumulative Gaussian function)

$$\text{Percent correct} = \frac{100}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\text{TMR}} \exp\left(\frac{-(\text{TMR}-\text{SRT})^2}{2\sigma^2}\right), \quad (1)$$

where SRT is the speech-reception threshold[3] (dB), σ is related to the slope of the function, and TMR is the target-to-masker ratio (dB). Figure 1 shows sample data from one listener, along with the best-fitting curve (heavy) according to Eq. (1). The other, lighter curves in the figure are the fits to the data from the other seven listeners in that experimental condition. The two-parameter model assumes that listeners' peak reception performance is 100%. This assumption may be valid for the 24- and 8-channel conditions, but it is probably not valid for the 4-channel condition. Therefore, the initial model had a third parameter, associated with the peak performance. However, the goodness of fit and the estimated SRTs of the three-parameter model were very similar to those of the two-parameter model, leading us to select the model with fewer parameters.

The two-parameter model provided generally good fits to the curves of performance as a function of TMR. Presented in Table II are the mean values of SRT, σ, and standard error of fit, averaged across listeners. The standard deviations are shown in parentheses. The individual standard errors of fit[4] had a mean of 7.25% with a standard deviation of 3.7% (median of 7.01% and a worst case of 20.93%). Combined according to experimental conditions, the average standard errors of fit (Table II) were generally less than 10%.

B. Speech-reception thresholds

In general, performance across listeners was reasonably consistent, so only the mean SRT values as a function of masker condition and processing condition are plotted in Fig.

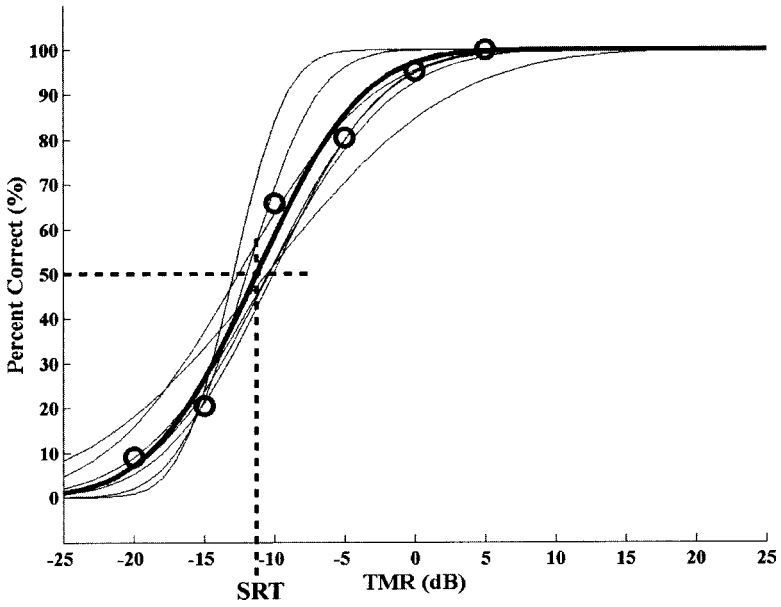


FIG. 1. An example of the two-parameter sigmoid model fitting procedure. The two-parameter sigmoid model (heavy line) is fitted to the speech reception performance data of an individual listener (open circles). The light lines are the functions fitted to the data of the other listeners in the same experimental condition (data not shown). The speech reception threshold (SRT) is the target-to-masker ratio (TMR), where 50% of the words were correctly identified.

2. The mean SRT values and standard errors of means (Table II) were derived from the SRT values of individual model fits. Since an SRT value is the TMR where 50% of the keywords are correctly identified, a higher SRT value implies a condition more detrimental to speech reception. Figure 2 shows that SRT values in all masker conditions were strongly affected by implant processing. As the number of spectral channels decreased, the SRT values under all masker types increased. However, the rate of increase differed between masker types. As the number of spectral channels decreased, SRT values increased faster in the presence of fluctuating maskers than in the presence of steady-state noise, particularly for the single-talker maskers.

A two-way mixed-design analysis of variance (ANOVA) was performed using STATISTICA (StatSoft, Tulsa, OK) to determine the statistical significance of the findings, with SRT as the dependent variable, processing condition as the between-subjects factor, and masker type as the within-subjects factor. The ANOVA indicated that both main factors and their interaction were statistically significant (processing condition: $F_{3,28}=218.1$; masker type: $F_{3,84}=7.5$; interaction: $F_{9,84}=8.7$; $p<0.001$ in all cases). A *post hoc* test according to Fisher's LSD (alpha=0.05) indicated several significant differences between the different experimental conditions, as outlined below.

TABLE II. Mean sigmoidal model parameter values [Eq. (1)], averaged across listeners. The standard deviations are in parentheses. The standard error of fit provides a numerical indicator for how well the model fits the data. SRT is the speech reception threshold, and σ is related to the slope of the function.

Processing condition	Masker type	SRT (dB TMR)	σ	Standard error of fit (%)
Unprocessed	Male interference	−10.3(2.4)	7.5(1.9)	8.3(3.5)
	Female interference	−11.3(1.1)	6.3(2.3)	6.8(3.2)
	Modulated noise	−9.1(0.6)	4.1(1.5)	5.6(2.3)
	Steady-state noise	−6.7(0.8)	3.4(0.7)	5.5(3.4)
24 channels	Male interference	0.7(1.8)	5.1(1.1)	6.4(3.4)
	Female interference	0.6(1.8)	5.0(1.0)	4.8(2.1)
	Modulated noise	−3.3(0.6)	5.0(1.1)	4.0(2.3)
	Steady-state noise	−1.2(1.2)	3.3(0.6)	3.8(1.8)
8 channels	Male interference	6.4(2.2)	6.0(2.0)	9.1(2.0)
	Female interference	6.7(1.5)	5.2(1.4)	7.2(1.7)
	Modulated noise	4.6(2.1)	7.7(1.5)	6.8(3.4)
	Steady-state noise	4.2(0.8)	4.1(1.2)	5.4(1.9)
4 channels	Male interference	18.1(3.1)	14.1(1.8)	9.9(3.1)
	Female interference	18.3(4.3)	15.3(3.3)	12.5(3.8)
	Modulated noise	15.6(4.4)	18.7(4.8)	9.3(2.2)
	Steady-state noise	14.9(5.4)	19.3(9.2)	10.6(3.4)

In the unprocessed conditions, the steady-state noise masker was significantly more effective than any of the modulated maskers. However, under implant processing the reverse was true, with the exception of the 24-channel processed modulated speech-shaped noise condition. These differential effects are illustrated in Fig. 3, which treats the steady-state masker as the baseline condition and plots the differences in SRT values between the steady-state noise and the other maskers as a function of processing condition. Significant differences between SRTs in steady-state noise and those in the other conditions are labeled with asterisks in Fig. 3.

The single-talker interferers produced significantly higher SRT values than steady-state noise in all processed conditions (i.e., 24, 8, and 4 channels). In contrast, the modulated noise produced lower SRT values than the steady-state noise in the 24-channel condition, and was not significantly different from the steady-state noise in the 8- and 4-channel conditions. As illustrated in Fig. 2, in all conditions the transition from unprocessed to 24-channel processing resulted in a large increase in SRT value. This is despite the fact that the 24-channel condition represents frequency resolution approaching that found in normal-hearing listeners. This finding is explored further in Sec. IV C.

There was no significant difference in the SRT values between the male and female single-talker maskers in any processing condition (Fig. 2). Given our hypothesized effect

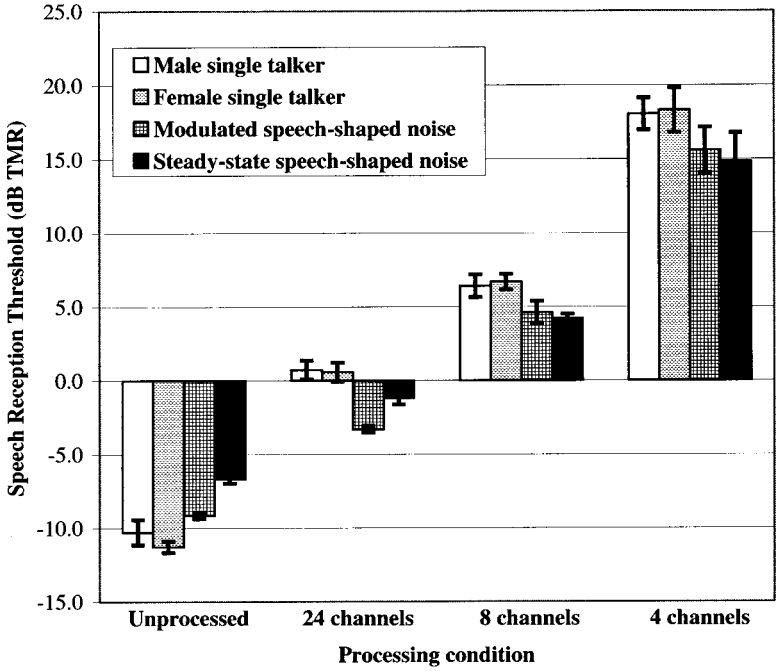


FIG. 2. Speech reception threshold, in terms of target-to-masker ratio, as a function of processing condition in the presence of a male single talker (unshaded), a female single talker (dotted), modulated noise (grid), and steady-state noise (solid). The plotted values and their respective standard deviations can be found in Table II.

of $F0$ differences in source segregation, this may seem unexpected. This finding is explored further in Sec. IV D.

IV. DISCUSSION

A. Single-talker versus steady-state noise interference

Our results in the unprocessed conditions are consistent with previous studies in showing that SRT values are lower for single-talker interferers than for steady-state noise (e.g., Festen and Plomp, 1990; Peissig and Kollmeier, 1997; Peters *et al.*, 1998). The improved performance found with single-talker interference, relative to steady-state noise, has been ascribed to listeners' ability to gain information from temporal or spectral minima in the maskers. However, to make use of local masker minima, the listener must have cues to distinguish the target from masker. Voice $F0$ is a generally accepted segregation cue for normal-hearing listeners (Brokx and Nooteboom, 1982; Bird and Darwin, 1998; Freyman *et al.*, 1999; Brungart, 2001). Our hypothesis was that the reduction in $F0$ cues, produced by simulated cochlear-implant processing, would particularly affect speech reception where the ability to discriminate the target from the masker is thought to play an important role. The results from the processed conditions are consistent with the hypothesis: not only are the benefits of spectral and temporal masker dips eliminated, but the single-talker interferers go from being the least effective maskers in the unprocessed conditions to being the most effective maskers in all the processed conditions (see Figs. 2 and 3).

B. Modulated noise versus steady-state noise

If simulated cochlear-implant processing led to a global inability to use temporal minima in fluctuating maskers, the

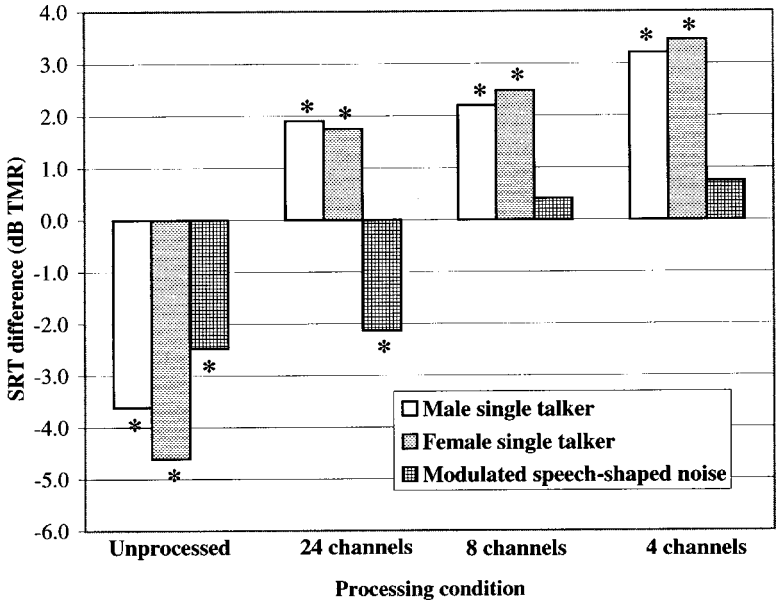


FIG. 3. SRT differences between the steady-state noise masker and the male single-talker (unshaded), female single-talker (dotted), and modulated-noise (grid) maskers are shown as a function of processing condition. Masked thresholds significantly different from those in the steady-state noise, according to Fisher's LSD test (alpha=0.05), are labeled by an asterisk.

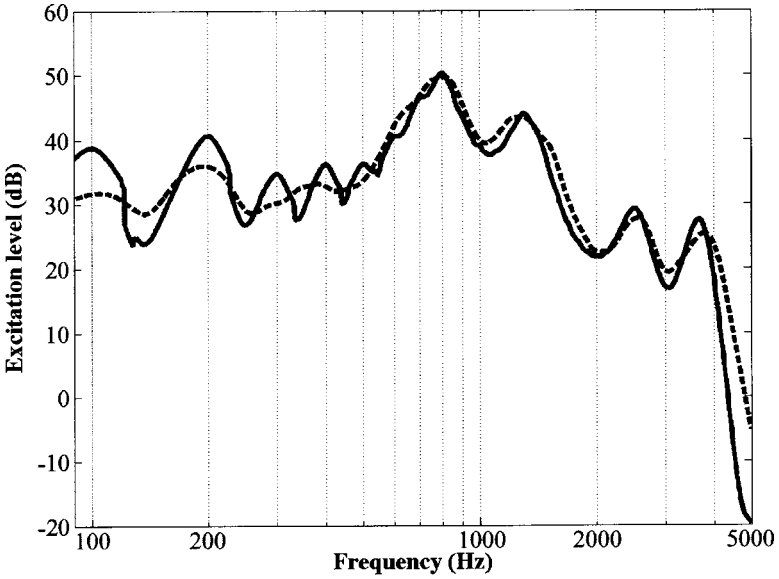


FIG. 4. An illustration of the difference in effective spectral resolution between unprocessed (solid) and 24-channel processed (dashed) conditions. This figure is the output of the excitation pattern model (Moore *et al.*, 1997) in response to a 500-ms Klatt synthesized vowel /a/, with an $F0$ of 100 Hz (Klatt, 1980).

same deterioration in performance would be expected in the modulated-noise masking conditions as was found for the single-talker interferers. In fact, the difference in performance between the modulated-noise and the steady-state-noise conditions remains roughly constant for the unprocessed and 24-channel conditions. Even for 8 and 4 channels, SRT values in the modulated-noise conditions are not significantly higher than in the steady-state noise conditions.

Without $F0$ cues, listeners may still maintain high levels of speech reception in the presence of interference by utilizing different cues. For example, when speech is presented in the presence of steady-state noise, a listener may be able to use common slowly varying envelope modulation as a cue for segregating the target from the noise, as most slow-varying envelope modulations will belong to the target. In the case of the amplitude-modulated speech-shaped noise masker, the noise is always modulated coherently across frequency. Speech, on the other hand, does not always modulate coherently across all frequencies. Listeners could use the more consistent comodulation of the amplitude-modulated noise as a cue for source segregation. However, to use comodulation as a segregation cue, spectral resolution must be sufficiently fine to distinguish the time-varying spectral changes of the target speech from the comodulated noise masker. This could account for the SRT difference between modulated speech-shaped noise and steady-state noise in the unprocessed and 24-channel processing conditions. If the spectral resolution is too coarse (e.g., in the 8- and 4-channel conditions) the stimulus representation of target speech will also exhibit very strong comodulation. This may eliminate differences in comodulation as a valid cue for distinguishing between masker and target, and may account for the lack of an SRT difference between the modulated speech-shaped noise and steady-state noise in the 8- and 4-channel processing conditions.

C. Unprocessed versus 24-channel processing

As shown in Fig. 2, performance with 24-channel processing was considerably worse than with no processing, for all masker types. This may seem surprising, given that the spectral resolution in the 24-channel processing condition was chosen to be similar to that found for normal-hearing listeners, with 3-dB filter bandwidths of 1.16 Cams. In Fig. 4, the excitation patterns (Moore *et al.*, 1997) for the vowel /a/ with and without 24-channel processing are plotted. It can be seen that the spectral peaks of the vowel are comparably well represented in both the processed (dashed) and unprocessed (solid) conditions. This may suggest that the temporal fine structure, discarded by the processing, while not necessary for speech recognition in quiet, may play an important role in segregating speech from interfering sounds. Similarly, it can be seen from Fig. 4 that the spectral resolution of the first few harmonics (below 500 Hz) is degraded in the 24-channel processing condition. While that information may not be important for speech reception *per se*, the first few harmonics carry important information about the stimulus $F0$. It is therefore possible that the loss of $F0$ information, due to a reduction in harmonic resolution and/or a loss of temporal fine-structure information, is responsible for the large difference in performance between the unprocessed and 24-channel conditions.

D. Male versus female single-talker interference

As mentioned in Sec. II, the motivation for using different gender single-talker interferers came from the observation that normal-hearing listeners benefit from $F0$ differences between target and interfering talkers. This benefit generally increases with increases in $F0$ differences (Assmann and Summerfield, 1990; 1994; Bird and Darwin, 1998). Our finding of no significant difference between the male and female interferers may therefore seem surprising. There are at least three possible explanations for this null effect.

The first possible explanation lies in the instantaneous $F0$ values. In many past studies of single-talker interference (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990; 1994; Bird and Darwin, 1998), the $F0$s of the targets and maskers were held constant, either through the use of

short-duration stimuli or synthesized speech with a fixed $F0$. In the present study, the single-talker interferers were taken from recorded books, where exaggerated prosody is common. Although the mean $F0$ of the male single-talker interference (111.4 Hz) was approximately equal to that of the target (110.8 Hz), and the mean $F0$ of the female single-talker interference (129.4 Hz) was about three semitones higher, the natural variations in $F0$ were left unaltered. As a result, the $F0$ differences between the target and single-talker interference were distributed such that the probability[5] of a two-semitone difference in $F0$ between the target and male single-talker interference was 0.69, and between the target and the female single-talker interference was 0.76. The lack of difference in SRT values between the male and female masker may therefore be due to the large differences in instantaneous $F0$ between the target and both maskers. However, contrary to the hypothesis, previous studies showed little or no improvements in identification as a result of time-varying $F0$s, as compared to constant $F0$s (Darwin and Culling, 1990; Summerfield and Culling, 1992; Assmann, 1999). Their findings suggest that the instantaneous difference in $F0$ between the competing talkers in this study may not be the main factor behind the lack of difference between the two single-talker maskers.

The second possible explanation lies in the atypical $F0$ range of the female voice used in this experiment. Adult female voices have an average $F0$ of around 220 Hz (Hillenbrand *et al.*, 1995). The finding of no significant difference between the male and female interferers in this study may be due to unusually low mean $F0$ (129.4 Hz) of the female interferer. The lack of a gender effect in this study may, therefore, not generalize to everyday situations.

The third possible explanation lies in the individual vocal characteristics of the talkers (e.g., vocal tract length, accent, speaking style, sentence level stress, etc.) The differences in vocal characteristics between the target and interfering talkers may have been sufficiently large to render any further improvement due to mean $F0$ difference negligible.

E. Importance of frequency selectivity and temporal fine structure: Implications for cochlear implants

Speech is an ecologically important stimulus for humans. If speech reception in quiet can be achieved with minimal spectral resolution and no temporal fine-structure information, then is the exquisite frequency selectivity and sensitivity to fine structure of the human auditory system necessary for speech communication? One important function of frequency selectivity may be found in earlier studies of spectral smearing (ter Keurs *et al.*, 1992; Baer and Moore, 1993) and of cochlear-implant simulations in steady-state noise (Dorman *et al.*, 1998; Fu *et al.*, 1998). From these studies, and from the present study, it can be seen that greater frequency selectivity results in lower signal-to-noise ratios necessary for speech reception. The present results suggest that sensitivity to low-frequency temporal fine structure (e.g., Meddis and O'Mard, 1997), or the spectral resolution of the lower harmonics (e.g., Terhardt, 1974) is critical to good speech reception in complex backgrounds. Performance is greatly affected by simulated implant processing, even with 24-channel resolution. The effect of stimulus processing is especially dramatic for the single-talker interferers, where an SRT benefit with respect to steady-state noise in the unprocessed condition is transformed into a deficit in all processed conditions. We hypothesize that the dramatic deterioration in performance is related to an inability to perceptually segregate the target from the masker, and that successful segregation relies on good frequency selectivity and $F0$ sensitivity.

The present results have possible implications for cochlear-implant design. As in previous studies (Dorman *et al.*, 1998; Fu *et al.*, 1998), the results support separating the spectrum into as many channels as possible, given the technical constraints of ensuring channel independence (Friesen *et al.*, 2001). However, the results also suggest that simply increasing the number of channels (at least to 24) may not assist in providing sufficient $F0$ information to successfully segregate a target from an acoustically complex background. The problem of presenting usable fine-structure information to implant users is current topic of research (e.g., Litvak *et al.*, 2001), and the present results provide further support for such endeavors. Finally, the large differences between performance in steady-state noise and performance in fluctuating backgrounds, particularly single-talker interferers, suggest that testing cochlear-implant patients in steady-state noise alone may underestimate the difficulties faced by such listeners in everyday acoustic environments.

V. SUMMARY

(1) Simulated cochlear-implant processing leads to a large deterioration in speech reception in the presence of a masker, even when the spectral resolution approaches that of normal hearing.

(2) Under simulated implant processing, single-talker interference is more detrimental to speech reception than steady-state noise. This is the converse of the situation found without processing, and it highlights the potential importance of frequency selectivity and temporal fine-structure information in segregating complex acoustic sources.

(3) In the presence of steady-state noise, the amplitude modulations associated with the target speech may provide useful source segregation cues, even under stimulated implant processing, provided that the spectral resolution is sufficiently fine.

(4) Using steady-state noise to test speech intelligibility may underestimate the difficulties experienced by cochlear-implant patients in everyday acoustic backgrounds.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant R01 DC 05216). We are grateful to Christophe Micheyl, Jeanie Krause, Peninah Rosengard, and Alan Asbeck for their comments on an earlier version of this manuscript. We also thank Peter Assmann, Bob Shannon, and John Culling for their many helpful suggestions.

[1]Shannon *et al.* (1995) processed speech by first dividing the audio spectrum into a small number of wide contiguous frequency bands. The temporal envelope of each band was extracted by half-wave rectification and low-pass filtering. The envelope derived from each band was then used to modulate separate white noises. The modulated noise from each envelope was frequency-limited by filtering with the same bandpass filters used in the original analysis band. The resulting modulated noises were then summed together.

[2]This is more frequently referred to as the ERB scale. However, as pointed out by Hartmann (1997), ERB simply refers to equivalent rectangular bandwidth, which could be used to define all estimates of auditory filter bandwidths. We, therefore, follow Hartmann's convention of referring to the scale proposed by Glasberg and Moore as the Cam scale, in recognition of its origins in the Cambridge laboratories. Described in Glasberg and Moore (1990), $\text{Cam} = 21.4 \log_{10}(0.004\,37 f + 1)$, where f is frequency in Hz.

[3]Speech reception threshold is the target-to-masker ratio (dB TMR) at which 50% of the words were correctly identified (see Fig. 1).

[4]The standard error of fit is the square root of the summed square of error divided by the residual degrees of freedom, $\sqrt{\text{SSE}/v}$, where $\text{SSE} = \Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2$. The residual degrees-of-freedom term (v) is defined as the number of response values (n) minus the number of fitted coefficients (m) estimated for the response values, $v = n - m$.

[5]The probability of a two-semitone difference in $F0$ was computed by integrating the $F0$ joint probability distribution function of the target and male or female single-talker interference.

Assmann, P. F. (**1999**). ''Fundamental frequency and the intelligibility of competing voices,'' Proc. 14th Int. Cong. of Phonetic Sci., pp. 179–182.

Assmann, P. F., and Summerfield, Q. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1994**). ''The contribution of waveform interactions to the perception of concurrent vowels,'' J. Acoust. Soc. Am. **95**, 471–484.

Baer, T., and Moore, B. C. J. (**1993**). ''Effects of spectral smearing on the intelligibility of sentences in the presence of noise,'' J. Acoust. Soc. Am. **94**, 1229–1241.

Bird, J., and Darwin, C. J. (**1998**). ''Effects of a difference in fundamental frequency in separating two sentences,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., and Clark, G. M. (**1984**). ''Speech processing strategies using an acoustic model of a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **76**, 104–110.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perceptual separation of simultaneous voices,'' J. Phonetics **10**, 23–36.

Brungart, D. S. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Burns, E. M., and Viemeister, N. F. (**1976**). ''Nonspectral pitch,'' J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). ''Played again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). ''Temporal pitch mechanisms in acoustic and electric hearing,'' J. Acoust. Soc. Am. **112**, 621–633.

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Darwin, C. J., and Culling, J. F. (**1990**). ''Speech perception seen through the ear,'' Speech Commun. **9**, 469–475.

de Cheveigné, A., and Kawahara, H. (**2002**). ''YIN, a fundamental frequency estimator for speech and music,'' J. Acoust. Soc. Am. **111**, 1917–1930.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). ''The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels,'' J. Acoust. Soc. Am. **104**, 3583–3585.

Drullman, R. (**1995**). ''Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level,'' J. Acoust. Soc. Am. **98**, 1796–1798.

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Fletcher, H., and Galt, R. H. (**1950**). ''The perception of speech and its relation to telephony,'' J. Acoust. Soc. Am. **22**, 89–151.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3588.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. J., Shannon, R. V., and Wang, X. S. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

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,'' Hear. Res. **47**, 103–138.

Glasberg, B. R., and Moore, B. C. J. (**1992**). ''Effects of envelope fluctuations on gap detection,'' Hear. Res. **64**, 81–92.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). ''Masking of speech by amplitude-modulated noise,'' J. Acoust. Soc. Am. **95**, 518–529.

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (Springer, New York).

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Houtgast, T., Steeneken, H. J. M., and Plomp, R. (**1980**). ''Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics,'' Acustica **46**, 60–72.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.

IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' IEEE Trans. Audio Electroacoust. **AU-17**(3), 225–246.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Litvak, L., Delgutte, B., and Eddington, D. (**2001**). ''Auditory nerve fiber responses to electrical stimulation: Modulated and unmodulated pulse trains,'' J. Acoust. Soc. Am. **110**, 368–379.

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

Miller, G. A., and Licklider, J. C. R. (**1950**). ''The intelligibility of interrupted speech,'' J. Acoust. Soc. Am. **22**, 167–173.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness, and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.

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

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).

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.

Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **101**, 1660–1670.

Peters, R., Moore, B., and Baer, T. (**1998**). ''Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people,'' J. Acoust. Soc. Am. **103**, 577–587.

Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). ''On the perceptual organization of speech,'' Psychol. Rev. **101**, 129–156.

Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency-modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M.

(**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Summerfield, Q., and Culling, J. F. (**1992**). ''Auditory segregation of competing voices: Absence of effects of FM or AM coherence,'' Philos. Trans. R. Soc. London, Ser. B **336**, 357–365.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). ''Effect of spectral envelope smearing on speech reception,'' J. Acoust. Soc. Am. **91**, 2872–2880.
Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.
Vliegen, J., and Oxenham, A. J. (**1999**). ''Sequential stream segregation in the absence of spectral cues,'' J. Acoust. Soc. Am. **105**, 339–346.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Modeling vocal fold motion with a hydrodynamic semicontinuum model

M. Drew LaMar
Department of Mathematics, University of Texas at Austin, Austin, Texas 78712

Yingyong Qi
Qualcomm Inc., 5775 Morehouse Drive, San Diego, California 92121

Jack Xin[a)]
Department of Mathematics and TICAM, University of Texas at Austin, Austin, Texas 78712

(Received 24 October 2001; accepted for publication 1 April 2003)

Vocal fold (VF) motion is a fundamental process in voice production, and is also a challenging problem for numerical computation because the VF dynamics depend on nonlinear coupling of air flow with the response of elastic channels (VF), which undergo opening and closing, and induce internal flow separation. The traditional modeling approach makes use of quasisteady flow approximation or Bernoulli's law which ignores air compressibility, and is known to be invalid during VF opening. A hydrodynamic semicontinuum system for VF motion is presented. The airflow is modeled by a modified quasi-one-dimensional Euler system with coupling to VF velocity. The VF is modeled by a lumped two mass system with a built-in geometric condition on flow separation. The modified Euler system contains the Bernoulli's law as a special case, and is derivable from the two-dimensional compressible Navier–Stokes equations in the inviscid limit. The computational domain contains also solid walls next to VFs (flexible walls). It is shown numerically that several salient features of VFs are captured, especially transients such as the double peaks of the driving subglottal pressures at the opening and the closing stages of VF motion consistent with fully resolved two-dimensional direct simulations, and experimental data. The system is much simpler to compute than a VF model based on two-dimensional Navier–Stokes system. [DOI: 10.1121/1.1577547]

PACS numbers: 43.70.Bk, 43.28.Ra, 43.28.Py, 43.40.Ga [AL]

I. INTRODUCTION

Vocal folds (VF) are the source of the human voice, and their motion is a fundamental process in speech production. Since VF motion is mechanical and results from the interaction of airflow and elastic response of VF, partial differential equations (PDEs) can be written down from classical continuum mechanics based on our knowledge of VF structures and air flow characteristics. A model of VF motion is naturally made of a certain form of compressible Navier–Stokes equations coupled with an elastic system on VF deformation.

In the past decade, much progress has been made in modeling the elastic aspect of VF. There are by now a hierarchy of elastic models for VF, from the two mass model of Ishizaka and Flanagan,[1] Bogaert,[2] to 16 mass as well as the continuum model of Titze and co-workers.[3–7] However, the modeling of airflow or the fluid aspect of VF is less explored. There are broadly two types of approaches in treating the glottal flow. One is to combine the Bernoulli's law in the bulk of the flow (quasisteady flow approximation) with either a quasi-steady pressure recovery theory,[1] or an analytical approximation downstream of the flow separation point.[2,3,8] Bernoulli's law oversimplifies the flow in the sense that partial derivatives in time are ignored, however, they are not negligible for transient effects such as pressure peaks. For example, it was realized[9] and concluded[10] that Bernoulli's law is not valid during one-fifth of the VF vibration cycle, especially at the VF opening and closure. The compressibility is ignored in the Bernoulli's law, especially in the subglottal region just before the VFs. The compressibility is essential for the pressure buildup to reopen the VFs.

The other approach is direct numerical simulation of channel flows and VF in the continuum. For example, Alipour *et al.*[6,7] simulated a two dimensional incompressible Navier–Stokes (NS) system and a finite element model of VF cover and body. Extensive computation and coding are expected to fully resolve the flows in the presence of moving boundaries, closures, and flow separation.

In this paper, we study an intermediate system consisting of a modified quasi-one-dimensional compressible Euler equations for the air flows and a recent version of the two mass model on VF.[2] The flow separation is accounted for using an empirical formula on the VF opening angle. The model is semicontinuum in that the flow variables are spatially continuous, while VFs are approximated by two masses (discrete). The modified quasi-one-dimensional Euler equations are extensions of those in the study of duct flows in aerodynamics,[11–14] with an additional coupling to the velocity of channel boundaries. This is the only viscous effect. It will be shown numerically that the modified Euler system is able to handle strong transient effects, such as rapid variation of subglottal pressures at VF opening and closing. Nu-

[a)]Author to whom correspondence should be addressed; electronic mail: jxin@math.utexas.edu

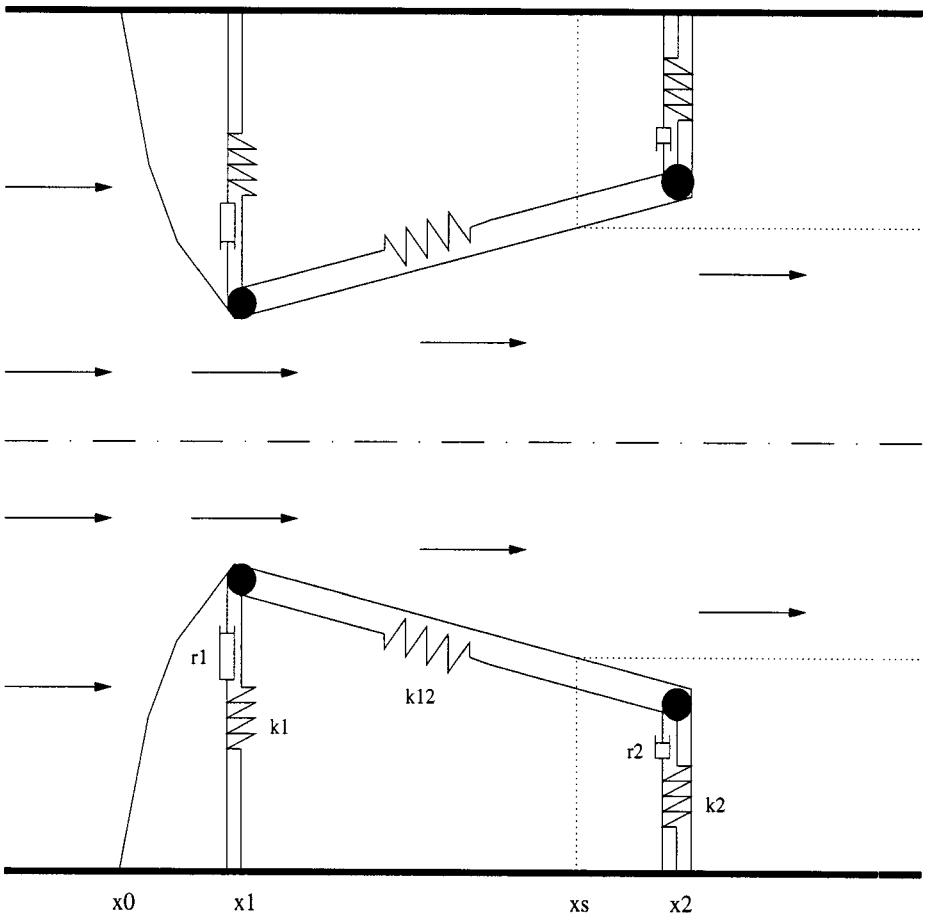


FIG. 1. A sketch of the flow and the two mass model. The computational domain is $[x_0,x_2]$. The channel width is fixed at x_0, and smoothly interpolated to mass one over $[x_0,x_1]$; x_s is the separation point for diverging glottis.

merical simulation of the model is less technical and many salient features of VF dynamics[6,8,10,15–17] are captured.

The rest of the paper is organized as follows. In Sec. II, the model equations are introduced, and related modeling issues addressed. In Sec. III, numerical method, convergence, and simulation results are discussed. It is shown numerically that model solutions recover several known VF characteristics, for example, the double VF inlet pressure peaks at VF opening and closure. The model robustness is shown by varying subglottal input pressure and plotting how air volume velocity changes as a function of time. The conclusion is in Sec. IV, followed by the acknowledgments. Appendix A contains a derivation of the modified quasi-one-dimensional Euler equations. Appendix B shows a linear stability analysis on the existence of oscillation modes near flat fold, to demonstrate the role of the boundary velocity coupling term in the modified Euler equations.

II. THE SEMICONTINUUM MODEL

Suppose the larynx is a two-dimensional channel with a finite mass elastic wall of cross section width $A(x,t)$. The VF is lumped into a sum of two masses connected by a spring, and each mass is connected to the solid wall by a spring and a damper, the common scenario in the two mass model,[1,2] see Fig. 1. The air flows from $x=x_0$ to $x=x_2$, and is modeled by the modified quasi-one-dimensional Euler system.

Conservation of mass,

$$(A\rho)_t+(\rho uA)_x=0, \tag{2.1}$$

ρ is the air density, u is the air velocity.

Reduced momentum equation,

$$(\rho uA)_t+(\rho u^2A)_x=-(pA)_x+A_xp+\rho uA_t, \tag{2.2}$$

p is the air pressure.

Assuming that the temperature is maintained as constant, the airflow is isothermal,[18] and the equation of state is

$$p=a^2\rho, \tag{2.3}$$

where a is the speed of sound. The cross section width A is a piecewise linear function in x determined by the displacements of the two masses (y_1,y_2), in the two-mass model system (Bogaert,[2] Ishizaka and Flanagan[1]):

$$m_1y_1''+r_1y_1'+k_1(y_1-y_{0,1})+k_{12}(y_1-y_2+y_{0,12})=F_1, \tag{2.4}$$

$$m_2y_2''+r_2y_2'+k_2(y_2-y_{0,2})+k_{12}(y_2-y_1-y_{0,12})=0, \tag{2.5}$$

where $F_1=L_g\int_{-L}^{x_s}p\,dx$, L_g the transverse (to the flow) dimension of VF, equal to 1 cm; y_i's are VF openings at locations x_i's, $i=1,2$; $x_s=x_2$ if there is no flow separation, and x_s=the location of flow separation if it occurs. The m_i, r_i, k_i, $i=1,2$, are mass density, damping and elastic spring constants. Mass one (lower mass) is situated near the VF entrance, and mass two (upper mass) is located towards the exit of the glottal region. Following Bogaert,[2] x_s will be estimated by an empirical formula on the degree of diver-

gence of the VF. Our complete VF model is the coupled system (2.1)–(2.5).

The viscous effect in the flow produces the term $\rho u A_t$ from the no-slip boundary condition of the two-dimensional flows, see derivation in Appendix A. Without this term, the above system is quasi-one-dimensional Euler in gas dynamics.[11–14] The extra term introduces coupling to wall velocity, and is critical in transferring energy from airflow into the VF, as the Titze theory[19] predicted. Appendix B shows that with this term, there exist oscillation modes near flat fold under a threshold condition similar to the one in Titze.[19] For simplicity, we ignored other viscous effects.

The two-mass model (2.4)–(2.5) is a recent improvement[2] of the original IF72[1] in that the flow separation point is not always at the VF exit, instead it depends on the glottal geometry. Flow separation basically refers to a change of flow behavior from being attached to the VF cover via a viscous boundary layer to a developed free jet with vortical structures and turbulent wake. Because of the vortical buildup, pressure near the wall is typically low, and can be approximated by setting it to zero (or ambient pressure) as done on mass two in (2.5). In converging glottis, there is no flow separation, however in diverging glottis, it occurs if the diverging angle is large enough. It is expedient for our modeling purpose to adopt a working hypothesis supported by experiments[2,8]

$$y_2/y_1 < 1.1 \Rightarrow x_s = x_2, \tag{2.6}$$

$$y_2/y_1 > 1.1 \Rightarrow x_s = x_1 + \frac{(x_2 - x_1)y_1}{10(y_2 - y_1)}, \quad y_s = 1.1 y_1. \tag{2.7}$$

Notice that the location of the flow separation is a variable depending on the diverging angle. It is worth pointing out that the assumptions made for deriving the reduced flow model are more accurate prior to the separation point. After the flow separation point, the reduced flow model needs to be properly corrected. For example, contribution of the viscous boundary layers can be introduced in the model to calculate the separation point more accurately than formula (2.6)–(2.7), and take into account the energy losses downstream of the separation point. A formulation of such a treatment with von Karman equations is given by Pelorson *et al.*[8] For simplicity of the model, we shall not pursue this task here, instead we rely on (2.6)–(2.7) as a simple way to incorporate separation effects. As the pressure after separation point is not contributing a force to the upper mass in (2.5), modeling error of ignoring viscous losses in the boundary layers is minimized within our model system. Viscous effects can be neglected upstream of the separation point for vocal flows.[8]

We also adopt the elastic collision (stopping) criterion[1,2] when the two sides of VF approach each other and close. When y_i's are smaller than a critical level y_c, then VF is considered closed, and (m_i, r_i, k_i) $(i=1,2)$ are adjusted to closure values.[1,2] In this case, the flow equations are solved only over $x \in [x_0, x_1]$, and in (2.4)–(2.5) the pressure force is adjusted to $F_1 = L_g \int_{x_0}^{x_1} p\, dx$. Due to constant input pressure p_0 at x_0, pressure at x_1 builds up. The two mass ODE's (ordinary differential equations) are still running even during VF closure, and in due time the increased pressure reopens VF.

The two-mass model by Bogaert[2] consists of (2.4)–(2.7) along with the above elastic collision criterion. In Bogaert,[2] the forcing term on mass one F_1 is calculated with the Bernoulli's law in the open glottis case. In the closed glottis case, air pressure is equal to the ambient pressure over $[x_0, x_s]$. The two mass parameters are as listed in the table.

Our VF model system is solved as an initial boundary value problem on $x \in [x_0, x_2]$. The initial conditions are $p(x,0) = 0.14\,\text{Pa}$, the ambient air pressure; $u(x,0) = 0$; $(y_1, y_2)(0) = (y_{0,1}, y_{0,2})$, the equilibrium two-mass position. The inlet boundary conditions are $(p,u)(x_0,t) = (p_0(t), u_0(t))$, where $p_0(t)$ is a smooth increasing function such that $p_0(0) = 0.14\,\text{Pa}$, $p_0(0.1) = 700\,\text{Pa}$; $u_0(t)$ is a smooth increasing function such that $u_0(0) = 0$, $u_0(0.1) = 0.114\,285a$, a is the speed of sound. The outlet boundary condition is $(p_x, u_x)(x_2, t) = 0$. The advantage of such Neumann-type boundary conditions is to help the flow to go out of the computational domain, and minimize numerical boundary artifacts.

The major difference between our model and that of Bogaert[2] is that we do not make quasisteady approximation on the flow variables, instead we integrate the time-dependent system (2.1)–(2.2). This turns out to be particularly important for capturing transients near closure and reopening stages of VF motion. It is helpful to put the system (2.1)–(2.2) into a rescaled form. Let $v = u/a$, a is the speed of sound. Then,

$$\begin{aligned} &\frac{1}{a}(Ap)_t + (pvA)_x = 0, \\ &\frac{1}{a}(pvA)_t + (pv^2A)_x = -(pA)_x + A_x p + pvA_t/a, \end{aligned} \tag{2.8}$$

where typically $v = u/a \approx 0.1$, the Mach number. If we use the convenient cm g ms unit, $a = 35\,\text{cm/ms}$, $1/a$ is a small parameter. If we ignore the terms with a, we have exactly Bernoulli's law for quasisteady flows. These seemingly small terms are essential especially during opening stage of VF, and should be kept for an accurate time-dependent solution.

III. NUMERICAL METHOD AND SIMULATION RESULTS

For given VF shape, $A(x,t)$, the flow system (2.1)–(2.2) is of the form

$$U_t + (F(U))_x = G(U), \tag{3.1}$$

the so-called conservation law (see Ref. 18 and references therein) with lower order source term G. The function F is the flux function. We implemented a first-order finite difference method, where time marching is split into two steps. In the first step $(t = nk \to (n + \frac{1}{2})k)$, we solve the conservation law $U_t + (F(U))_x = 0$ with explicit Lax–Friedrichs method,[18]

$$U_j^{n+1/2} = \frac{1}{2}(U_{j-1}^n + U_{j+1}^n) - \frac{k}{2h}(F(U_{j+1}^n) - F(U_{j-1}^n)), \tag{3.2}$$

TABLE I. Two mass model parameters in cgs unit.

m_1	0.17 g
m_2	0.03 g
x_2-x_1	0.2 cm
x_1-x_0	0.05 cm
$k_{1,\text{open}}$	45 kdynes
$k_{1,\text{closed}}$	180 kdynes
$y_{0,1}$	0 cm
$k_{2,\text{open}}$	8 kdynes
$k_{2,\text{closed}}$	32 kdynes
$y_{0,2}$	0.0 cm
k_{12}	25 kdynes
$y_{0,12}$	0 cm
y_c	0.001 cm
$A(x_0,t)$	2 cm
$r_{1,\text{open}}$	17.5 dynes/(cm s)
$r_{1,\text{closed}}$	192.4 dynes/(cm s)
$r_{2,\text{open}}$	18.6 dynes/(cm s)
$r_{2,\text{closed}}$	49.6 dynes/(cm s)

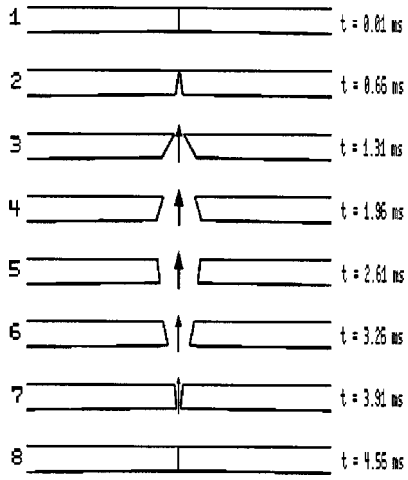


FIG. 3. A simulated VF vibration cycle, similar to the figure on p. 113 of Sataloff (Ref. 21).

where k and h are time step and spatial grid size. Here k must be small enough to ensure stability of the difference scheme and to keep the computed flow velocity positive (no back flow is allowed). In step two ($(n+\frac{1}{2})k \to (n+1)k$), we update the solution from $U^{n+\frac{1}{2}}$ to U^{n+1} by implicitly integrating ODEs: $U_t = G(U)$ in the flow equations, and the two-mass equations (2.4)–(2.5); where we apply central differencing in space and backward differencing in time. In the first step, U is updated using VF shape A at time $t = nk$; in the second step, the ODEs of the two-mass system and source terms are solved to update solutions to $(n+1)k$. We point out that when VF approach closure, the ODE's in step two become rather stiff, and this is the main reason to use implicit backward differencing in time.[20]

The numerical parameters used in our computation are space grid size $h = 0.25/(nx-1)$, where the computational domain is $x_2 - x_0 = 0.25$ cm, nx the total number of spatial grid points, $nx \geqslant 80$ ($h \leqslant 0.003\,165$); time step $k \leqslant 10^{-7}$. The time unit is ms, length unit cm, speed of sound $a = 35$ cm/ms. The two mass model parameters are listed in Table I. A convergence test of numerics is shown in the plot of the air volume velocity passing through the glottis as a function of time in Fig. 2. The numerical grids $h = 0.001\,572\,3$ ($nx = 160$), $k = dt = 2.5 \times 10^{-8}$ are used for the rest of the runs.

Now we describe our numerical results, and compare with figures in the literature either from experimental measurements or model calculations. In Fig. 3, we show a cycle of VF vibration, which resembles well the figure on p. 113 of Sataloff's *Scientific American* article.[21]

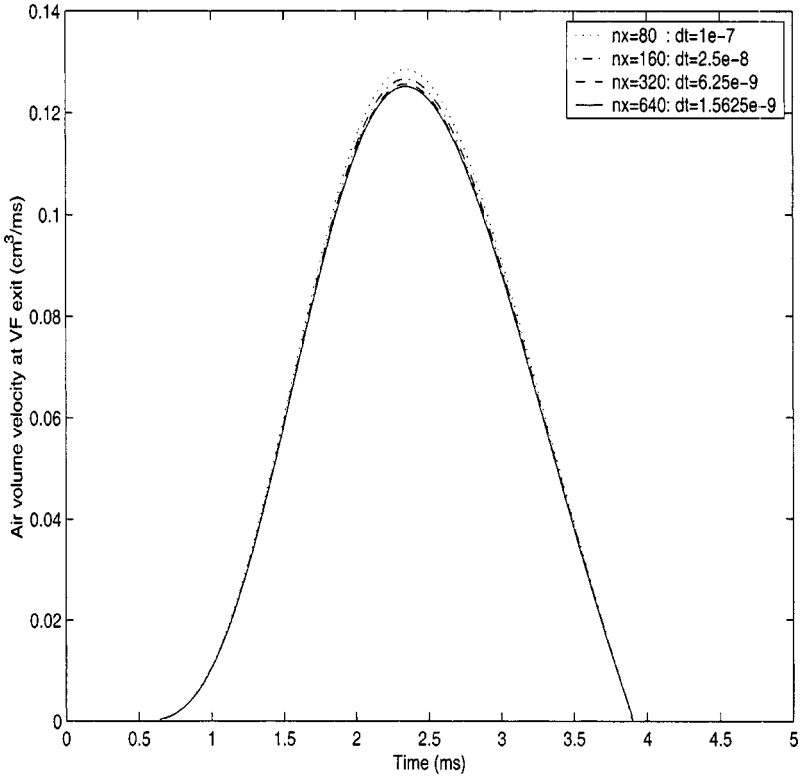


FIG. 2. The air volume velocity at VF exit x_2 to show convergence of numerical solutions under grid refinement.

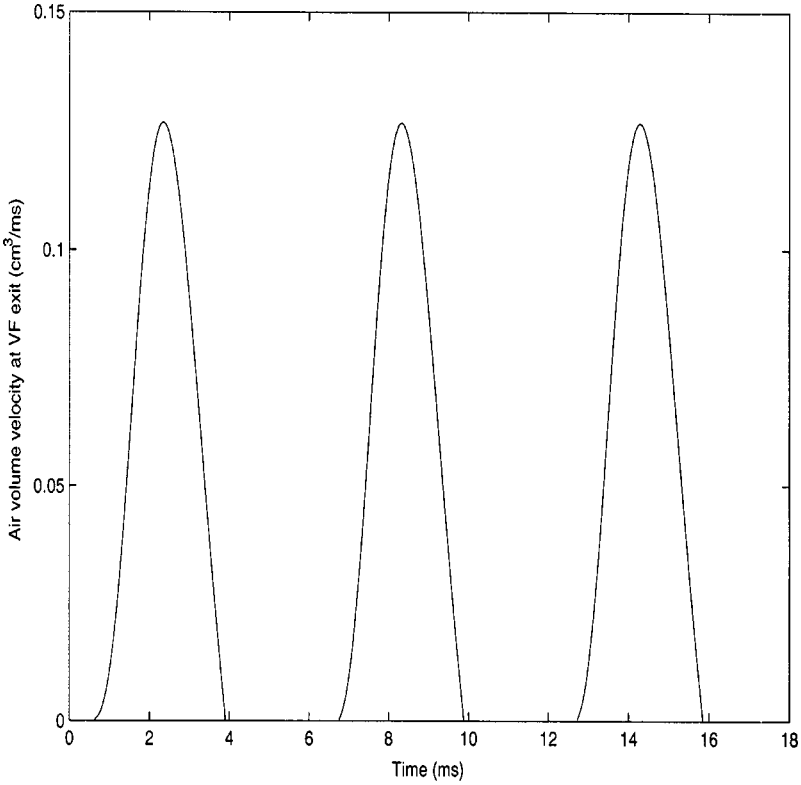


FIG. 4. Simulated VF air volume velocity (cm^3/ms) vs time at exit of VF from model (2.1)–(2.5).

In Fig. 4, we show the computed air volume velocity at the exit of VF, similar to Fig. 6, Fig. 7, Fig. 8 of Story and Titze,[3] and Fig. 6 of Alipour and Scherer.[6] Note that the pressure recovery downstream of the separation point is considered in the former,[3] and that a 2D incompressible NS calculation is used on a numerical domain covering a considerable wake flow region beyond x_2 in the latter[6] (see Fig. 4, p. 474). This comparison lends indirect support to the efficiency of our model treatment of VF flows.

Figure 5 is the experimentally measured intraglottal pressure on an excised canine larynx from Titze[15] (see also Jiang and Titze[16]), which showed the double peak (intraglottal) pressure structure, respectively, at VF opening and closing. Figure 6 is our computed subglottal pressure before mass one. The double peaks are present and resemble those

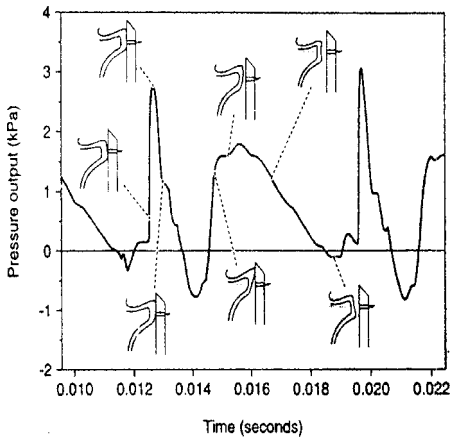


FIG. 5. Experimentally measured intraglottal pressure on excised canine larynx, reproduced Fig. 8 on p. 426 of Titze (Ref. 11) (with permission of the author and the publisher).

in Fig. 5, except that our second peak has a similar width as the first peak. Two main factors contribute to the difference. One is that the experiments have measured contact pressure when VFs are closed, and air pressure when VFs are open. Their intraglottal pressure has both components, while our computed subglottal pressure is only air pressure. As a result, the second peak in experiments due to air-pressure's gradual change is wider than the first which is mainly contact pressure. The other is that the closure treatment of two mass model differs from the actual VF closure. Our subglottal pressure is also in qualitative agreement with the computed subglottal air pressure in Fig. 5 (bottom frame),[6] which showed two peaks of nearly equal widths as well. Double peaks of intraglottal pressures have been computed[3] with a three-mass body-cover model and considerations of experimental conditions.

We also tested our model robustness under input pressure variation. In Fig. 7, we show a plot of air volume velocity vs time at VF exit for three subglottal pressures at x_0: 700 Pa, 1400 Pa, 2100 Pa with other parameters the same. We see that as subglottal pressures increase with other parameters fixed, air volume velocity curves get higher (at the peaks) and steeper (at the two sides). This agrees very well with Fig. 2.14(a), p. 78, of Stevens,[17] and is another support for our model.

We extend the flow domain to 0.5 cm downstream of x_2 (Fig. 1). Figure 8 shows the transglottal pressure as a function of time, calculated as the absolute difference between the instantaneous subglottal and supraglottal pressure values. The first and second peaks correspond to the opening of the glottis at mass 1 and 2, respectively, while the third peak corresponds to the closing of the glottis. The figure is similar to Fig. 5 in Alipour and Scherer,[6] as well as Fig. 9 of Austin

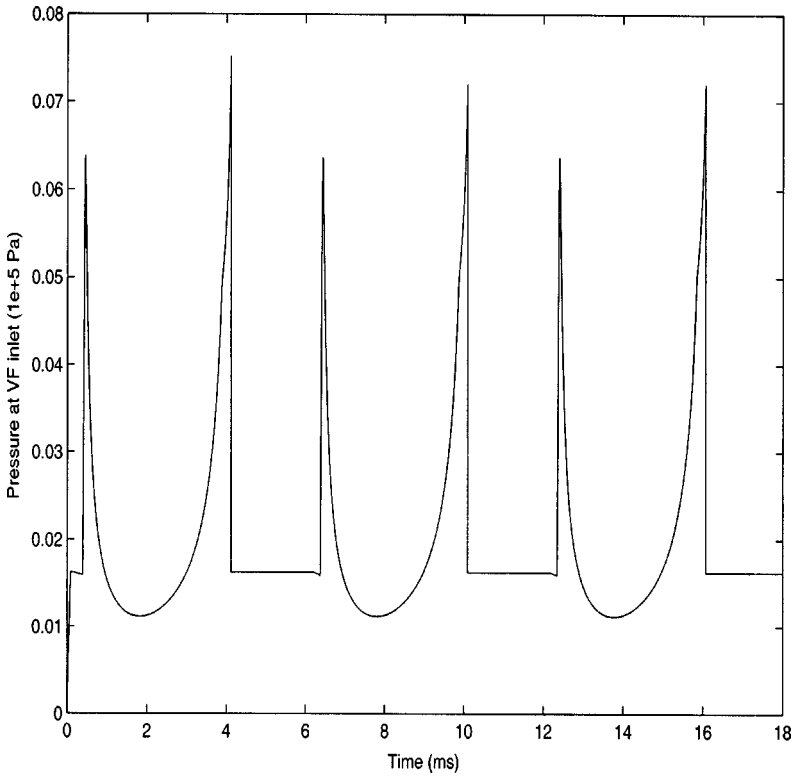


FIG. 6. The air pressure at VF inlet (before lower mass) versus time.

and Titze,[22] which displays experimental data collected from human subjects.

IV. CONCLUDING REMARKS

In this paper, we introduced and computed a semicontinuum VF model consisting of a modified quasi-one-dimensional Euler system and a recent two-mass model.[2] The flow part of the model is more accurate than a traditional treatment with Bernoulli's law, and also much simpler than a full two-dimensional Navier–Stokes system. We demonstrated numerical convergence and that the model solutions are in qualitative agreement with known VF characteristics.

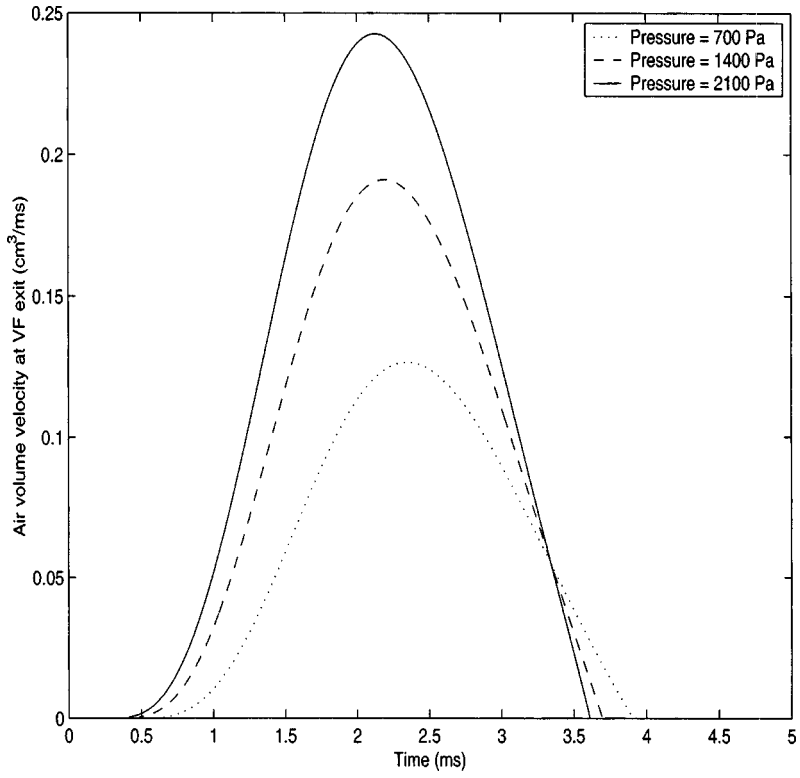


FIG. 7. The air volume velocity at x_2 for three values of the input subglottal pressures at x_0.

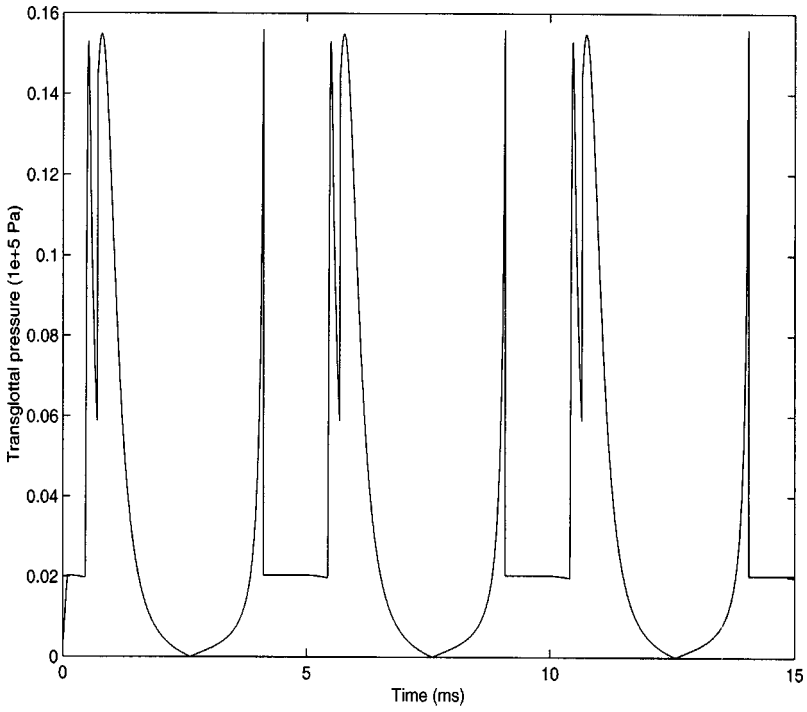


FIG. 8. Transglottal pressure through an extended domain, 0.5 cm past x_2 in Fig. 1.

ACKNOWLEDGMENTS

The authors wish to thank Professor I. Titze and Professor F. Alipour for helpful conversations and electronic mail communication on VF modeling, and for their recent work.[6] The authors would like to thank Dr. J. M. Hyman for suggesting Refs. 14 and 21, and Professor J. Keller for his helpful reading of an early version of the paper and his suggestions. M.D.L. thanks the UT continuing graduate fellowship, and Y.Q. thanks TICAM for hosting his visit at UT. This work was partially supported by ARO Grant No. DAAD 19-00-1-0524.

APPENDIX A: DERIVATION OF MODIFIED EULER SYSTEM

We derive the fluid part of the model system assuming that the fold varies in space and time as $A=A(x,t)$. Consider a two-dimensional slightly viscous subsonic air flow in a channel with spatially temporally varying cross section in two space dimensions, $\Omega_0=\Omega_0(t)=\{(x,y):x\in[-L,L], y\in[-A(x,t)/2,A(x,t)/2]\}$, where $A(x,t)$ denotes the channel width, or cross sectional area as the third dimension is uniform. The two-dimensional Navier–Stokes equations in differential form are (Batchelor,[23] p. 147) as follows.

Conservation of mass,

$$\rho_t+\nabla\cdot(\rho\vec{u})=0. \tag{A1}$$

Conservation of momentum,

$$(\rho\vec{u})_t=-\nabla\cdot(\rho(\vec{u}\otimes\vec{u}))+\mathrm{div}(\sigma\cdot\vec{n}), \tag{A2}$$

where σ is the stress tensor, $\sigma=(\sigma_{ij})=-p\,\delta_{ij}+d_{ij}$, and

$$d_{ij}=2\mu\left(e_{ij}-\frac{\mathrm{div}\,\vec{u}}{3}\,\delta_{ij}\right),$$

$$e_{ij}=\tfrac{1}{2}(u_{i,x_j}+u_{j,x_i}),$$

$$(x_1,x_2)\equiv(x,y),$$

μ is the fluid viscosity, $\Omega(t)$ is any volume element of the form $[\vec{u}=(u_1,u_2)]$

$$\Omega(t)=\{(x,y):x\in[a,b]\subset[-L,L],\quad y\in[-A(x,t)/2,A(x,t)/2]\}. \tag{A3}$$

The equation of state is either polytropic or isothermal.

The boundary conditions on $(\rho,\vec{u})$ are as follows:

(1) on the upper and lower boundaries $y=\pm A(x,t)/2$, $\rho_y=0$, and $\vec{u}=(0,\pm A_t/2)$, the velocity no slip boundary condition;
(2) at the inlet, $x=-L$, $p=p_0$, given subglottal pressure $(u_1,u_2)=(u_{1,0},u_{2,0})$, given input flow velocity. At the exit $(p,u_1,u_2)_x=0$, to help the waves go out of the domain freely.

We are only concerned with flows that are symmetric in the vertical. For positive but small viscosity, the flows are laminar in the interior of Ω_0 and form viscous boundary layers near the upper and lower edges. The vertically averaged flow quantities are expected to be much less influenced by the boundary layer behavior as long as $A(x,t)$ is much larger than $O(\mu^{1/2})$.

Let us assume that the flow variables obey

$$\begin{aligned}
&|u_{1,y}|\ll|u_{1,x}|,\\
&|u_{2,y}|\ll|u_{1,x}|, \quad \text{away from boundaries of } \Omega_0,\\
&|\vec{u}_y|\gg|\vec{u}_x|, \quad \text{near the boundaries of } \Omega_0,\\
&|\rho_y|\ll|\rho_x|, \quad \text{throughout } \Omega_0.
\end{aligned} \tag{A4}$$

These are consistent with physical observations in the viscous boundary layers (Batchelor,[23] p. 302), namely, there are large vertical velocity gradients, yet small pressure or density gradients in the boundary layers. The boundary layers are of width $O(\mu^{1/2})$. Denote by $\overline{\rho}$, $\overline{u_1}$, the vertical averages of ρ and u_1. Note that the exterior normal $\vec{n}=(-A_x/2,1)/(1+A_x^2/4)^{1/2}$ if $y=A/2$, $\vec{n}=(-A_x/2,-1)/(1+A_x^2/4)^{1/2}$ if $y=-A/2$.

Let $a=x$, $b=x+\delta x$, $\delta x \ll 1$, t slightly larger than t_0. We have

$$\frac{d}{dt}\int_{\Omega(t)}\rho\, dV=\frac{d}{dt}\int_{\Omega(t_0)}\rho J(t)dV$$

$$=\int_{\Omega(t_0)}\rho_t J(t)dV+\int_{\Omega(t_0)}\rho J_t\, dV, \tag{A5}$$

where $J(t)$ is the Jacobian of volume change from a reference time t_0 to t. Since $\Omega(t)$ is now a thin slice, $J(t)=A(t)/A(t_0)$ for small δx, and $J_t=A_t(t)/A(t_0)$. The second integral in (A5) is

$$\int_{\Omega(t_0)}\rho J_t\, dV=\overline{\rho}\frac{A_t(t)}{A(t_0)}A(t_0)\,\delta x=\overline{\rho}A_t(t)\,\delta x. \tag{A6}$$

The first integral is simplified using (A1) as

$$\int_{\Omega(t_0)}\rho_t J(t)dV=\int_{\Omega(t)}\rho_t\, dV=-\int_{\partial\Omega(t)}\rho\vec{u}\cdot\vec{n}\, dS. \tag{A7}$$

We calculate the last integral of (A7) further as follows:

$$\begin{aligned}\int_{\partial\Omega}\rho\vec{u}\cdot\vec{n}\, ds=&\int_{-A/2}^{A/2}(-\rho u_1)(x,y,t)dy+\int_{-A/2}^{A/2}(\rho u_1)\\&\times(x+\delta x,y,t)dy+\int_x^{x+\delta x}\rho\cdot(0,A_t/2)\\&\cdot(-A_x/2,1)dx+\int_x^{x+\delta x}\rho\\&\cdot(0,-A_t/2)\cdot(-A_x/2,-1)dx\\=&\overline{\rho u_1}A|_x^{x+\delta x}+\frac{\delta x}{2}(\rho A_t)|_{y=A/2}\\&+\frac{\delta x}{2}(\rho A_t)|_{y=-A/2}+O((\delta x)^2)\\\approx&(\overline{\rho}\cdot\overline{u_1}A)|_x^{x+\delta x}+\overline{\rho}A_t\delta x+O((\delta x)^2),\end{aligned} \tag{A8}$$

where we have used the smallness of ρ_y to approximate $\rho|_{y=\pm A/2}$ by $\overline{\rho}$ and $\overline{\rho u_1}$ by $\overline{\rho}\cdot\overline{u_1}$. Combining (A5)–(A7), (A8) with

$$\frac{d}{dt}\int_\Omega\rho\, dV=(\overline{\rho}A\,\delta x)_t+O((\delta x)^2), \tag{A9}$$

dividing by δx and sending it to zero, we have

$$(\overline{\rho}A)_t+(\overline{\rho}\cdot\overline{u_1}A)_x=0,$$

which is (2.1).

Next consider $i=1$ in the momentum equation, $a=x$, $b=x+\delta x$. We have similarly with (2.6),

$$\begin{aligned}\frac{d}{dt}\int_{\Omega(t)}\rho u_1\, dV=&\int_{\Omega(t)}(\rho u_1)_t\, dV+\int_{\Omega(t_0)}\rho u_1 J_t\, dV\\=&-\int_{\partial\Omega(t)}\rho u_1\vec{u}\cdot\vec{n}\, dS+\int_{\partial\Omega(t)}\sigma_{1,j}\cdot\vec{n}_j\, dS\\&+\overline{\rho u_1}A_t\delta x+O(\delta x^2).\end{aligned} \tag{A10}$$

We calculate the integrals of (A10) as

$$\begin{aligned}\frac{d}{dt}\int_\Omega\rho u_1\, dV=&(\overline{\rho u_1}A)_t\delta x+O((\delta x)^2)\\\approx&(\overline{\rho}\cdot\overline{u}A)_t\cdot\delta x+O((\delta x)^2).\end{aligned} \tag{A11}$$

Using $u_1=0$ on the upper and lower boundaries, a similar calculation as (A8) gives

$$\int_{\partial\Omega}\rho u_1\vec{u}\cdot\vec{n}\, dS=(\overline{\rho}\cdot\overline{u_1^2}A)|_x^{x+\delta x}+O(\delta x\mu^{1/2}), \tag{A12}$$

where the smallness of $u_{1,y}$ in the interior and small width of boundary layer $O(\mu^{1/2})$ gives the $O(\mu^{1/2})$ for approximating $\overline{u_1^2}$ by $\overline{u_1}\cdot\overline{u_1}$.

Remark 6.1: Notice that for inviscid flows, we would have an additional term $\overline{\rho u_1}A_t\delta x$, which would cancel the third term on the right-hand side of (A10). As a result, the $A_t u/A$ term would be absent from the momentum equation (2.2).

Let us continue to calculate

$$\begin{aligned}\int_{\partial\Omega}-p\,\delta_{1,j}n_j\, dS\approx&-\overline{p}A|_x^{x+\delta x}+\int_x^{x+\delta x}pA_x\, dx\\=&-\overline{p}A|_x^{x+\delta x}+\overline{p}A_x\delta x+O((\delta x)^2).\end{aligned}$$

Noticing that

$$d_{11}=2\mu(u_{1,x}-(u_{1,x}+u_{2,y})/3),\quad d_{12}=2\mu(u_{1,y}+u_{2,x}).$$

It follows that

$$\overline{d_{11}}=\frac{4}{3}\mu\overline{u}_{1,x}-\frac{2\mu A_t}{3A}.$$

Thus the contribution from the left and right boundaries located at x and $x+\delta x$ is

$$\sum_{l,r}\int_{l,r}d_{11}n_1=A\overline{d_{11}}|_x^{x+\delta x}=\frac{4}{3}A\mu\overline{u}_{1,x}|_x^{x+\delta x}-\left.\frac{2\mu A_t}{3}\right|_x^{x+\delta x}. \tag{A13}$$

The contribution from the upper and lower boundaries is

$$\begin{aligned}\sum_{\pm}\int_{y=\pm A/2}d_{11}n_1\, dS=&-d_{11}A_x\delta x/2|_{y=A/2}\\&-d_{11}A_x\delta x/2|_{y=-A/2}\\=&\mu\,\delta x\sum_{\pm}O(\partial_y\vec{u})|_{y=\pm A/2}.\end{aligned} \tag{A14}$$

Similarly,

$$\sum_{\pm}\int_{y=\pm A/2}d_{12}n_2\, dS=\mu\,\delta x\sum_{\pm}O(\partial_y\vec{u})|_{y=\pm A/2}. \tag{A15}$$

Since $\partial_y \vec{u}|_{y=\pm A/2} = O(\mu^{-1/2})$, the viscous flux from the boundary layers are $O(\delta x \mu^{1/2})$, much larger than the averaged viscous term $\delta x(4\mu/3)(A\overline{u}_{1x})_x = O(\delta x \mu)$. We notice that the vertically averaged quantities have little dependence on the viscous boundary layers unless A is on the order $O(\mu^{1/2})$. Hence the quantities from upper and lower edges in (A14) and (A15), and that in (A12), should balance themselves. Omitting them altogether, and combining remaining terms that involve only $\overline{u}_1$, $\overline{\rho}$ in the bulk, we end up with (after dividing by δx and sending it to zero)

$$(\overline{\rho}\cdot\overline{u}_1 A)_t + (\overline{\rho}\cdot\overline{u}_1^2 A)_x = -(\overline{p}A)_x + A_x\overline{p} + \overline{\rho}\,\overline{u}_1 A_t + \frac{4\mu}{3}(A\overline{u}_{1x})_x - 2\mu A_{tx}/3, \qquad \text{(A16)}$$

which gives the modified Euler (2.2) in the inviscid limit $\mu \to 0$.

APPENDIX B: OSCILLATION MODES NEAR FLAT FOLD

We present a linear analysis of oscillation modes when flow system (2.1)–(2.3) is coupled with a continuous wave model of VF instead of two-mass model. Our objective is to simplify the calculation and understand the VF velocity coupling term $\rho u A_t$. The wave equation is

$$m y_{tt} = \sigma y_{xx} - \alpha y_t - \beta y + p + f_m, \qquad \text{(B1)}$$

where σ represents the longitudinal tension, m mass density, α the damping, and β the stiffness. The pressure p acts pointwise on the fold $y = y(t,x) = A(t,x)/2$. The forcing f_m maintains an equilibrium position.

Let us consider the constant (equilibrium) state $(\rho,u,p,y) = (\rho_0, u_0, p_0, y_0)$, $u_0 > 0$, such that

$$-\beta y_0 + p_0 = f_m, \quad p_0 = a^2\rho_0, \qquad \text{(B2)}$$

and a nearby state $u = u_0 + \hat{u}$, $p = p_0 + \hat{p}$, $\rho = \rho_0 + \hat{\rho}$, $y = y_0 + \hat{y}$, where the hat variables are small perturbations. The linearized system is

$$(y_0\hat{\rho} + \hat{y}\rho_0)_t + (\rho_0 y_0 \hat{u} + \rho_0 u_0 \hat{y} + u_0 y_0 \hat{\rho})_x = 0, \qquad \text{(B3)}$$

$$\hat{u}_t + u_0\hat{u}_x + \frac{1}{\rho_0}\hat{p}_x = \frac{u_0}{y_0}\hat{y}_t, \qquad \text{(B4)}$$

$$m\hat{y}_{tt} = \sigma\hat{y}_{xx} - \alpha\hat{y}_t - \beta\hat{y} + \hat{p}. \qquad \text{(B5)}$$

Eliminating the $(\hat{\rho},\hat{u},\hat{p})$ variables using (B3) and (B4) and the equation of state $p = a^2\rho$, we obtain (denoting $\Gamma \equiv a^{-2}$):

$$y_0[\Gamma\partial_{tt} + 2\Gamma u_0\partial_{xt} + (\Gamma u_0^2 - 1)\partial_{xx}]\cdot(\hat{y}_{tt} - \sigma\hat{y}_{xx} + \alpha\hat{y}_t + \beta\hat{y}) = -\rho_0[(\partial_t + u_0\partial_x)^2\hat{y} + u_0\hat{y}_t]. \qquad \text{(B6)}$$

Seek a mode solution of the form $\hat{y} = e^{imx+\lambda t}$, m real, we end up with the following algebraic equation of degree four for λ:

$$[\Gamma\lambda^2 + 2\Gamma u_0 m\lambda i + (1 - \Gamma u_0^2)m^2]\cdot[\lambda^2 + \sigma m^2 + \alpha\lambda + \beta] = -\rho_0 y_0^{-1}[\lambda^2 + 2u_0 im\lambda - u_0^2 m^2] - \frac{u_0\rho_0\lambda}{y_0}, \qquad \text{(B7)}$$

or

$$\Gamma\lambda^4 + (\alpha\Gamma + 2\Gamma u_0 mi)\lambda^3 + (\Gamma(\beta + \sigma m^2) + 2\alpha\Gamma u_0 mi + \rho_0 y_0^{-1} + (1 - \Gamma u_0^2)m^2)\lambda^2 + \left(2\Gamma u_0 mi(\beta + \sigma m^2) - \alpha m^2(\Gamma u_0^2 - 1) + \rho_0 y_0^{-1}(2u_0 im) + \frac{u_0\rho_0}{y_0}\right)\lambda + [(\beta + \sigma m^2)(1 - \Gamma u_0^2)m^2 + \rho_0 y_0^{-1}(-u_0^2 m^2)] = 0. \qquad \text{(B8)}$$

We show the following proposition.

Proposition 7.1: If

$$\rho_0 u_0^2 > \alpha(\rho_0 u_0 + \Gamma\beta u_0 y_0), \qquad \text{(B9)}$$

(B8) has a pair of pure imaginary solution $\lambda = \pm i\eta$, $\eta \neq 0$ real, implying the existence of a pair of oscillatory modes to the linearized system (B4) and (B5) of the form $e^{\pm i(mx+\eta t)}$, for real and nonzero numbers m and η.

Proof: Let $\lambda = i\eta$ in (B8), where η is real. The real and imaginary parts give, respectively,

$$\Gamma\eta^4 + 2\Gamma u_0 m\eta^3 - \left[\Gamma(\sigma m^2 + \beta) + m^2(1 - \Gamma u_0^2) + \frac{\rho_0}{y_0}\right]\eta^2 + \left[-2\Gamma u_0 m(\beta + \sigma m^2) - \frac{2\rho_0 u_0 m}{y_0}\right]\eta + \left[m^2(1 - \Gamma u_0^2)(\sigma m^2 + \beta) - \frac{2u_0^2 m^2\rho_0}{y_0}\right] = 0 \qquad \text{(B10)}$$

and

$$-\alpha\Gamma\eta^3 - 2\alpha\Gamma u_0 m\eta^2 + m^2\alpha(1 - \Gamma u_0^2)\eta + \frac{u_0\rho_0\eta}{y_0} = 0. \qquad \text{(B11)}$$

For $\eta \neq 0$, $\alpha \neq 0$, we have from (B11),

$$\Gamma\eta^2 + 2\Gamma u_0 m\eta - m^2(1 - \Gamma u_0^2) - \frac{u_0\rho_0}{y_0\alpha} = 0,$$

so

$$\eta = -u_0 m \pm a\sqrt{m^2 + u_0\rho_0/(y_0\alpha)}, \qquad \text{(B12)}$$

which is not equal to zero in case of minus sign ($u_0 > 0$). Now we regard the left-hand side of (B11) as a continuous function of m, call it $F(m)$. For $|m| \gg 1$, $\eta \sim (-u_0 - a)m$, direct calculation shows

$$F(m) \sim -\frac{\rho_0}{\Gamma y_0} m^2 < 0.$$

While for $|m| \ll 1$,

$$\eta \sim -\sqrt{\frac{u_0\rho_0}{\Gamma y_0\alpha}} + O(m),$$

$$F(m) \sim \frac{u_0\rho_0}{\Gamma y_0\alpha}\left(\frac{u_0\rho_0}{y_0\alpha} - \Gamma\beta - \frac{\rho_0}{y_0}\right) > 0,$$

provided

$$\frac{u_0\rho_0}{y_0} > \alpha\left(\frac{\rho_0}{y_0} + \Gamma\beta\right), \qquad \text{(B13)}$$

holds, which is just (B9). Under (B9), $F(m)=0$ has a nonzero real solution, hence an oscillatory mode solution exists to (B10) and (B11). Finally, noticing that equations (B10) and (B11) are invariant under the symmetry transform $(\eta,m)\rightarrow(-\eta,-m)$, we conclude that the oscillatory modes exist as a pair.

Remark 7.1: Condition (B9) says that the fluid energy must be large enough to overcome the fold damping due to α. Without the term $\rho u A_t$ in (2.2), the same calculation would show that $F(m)=-\rho_0 m^2/(\Gamma y_0)<0$ for all m, implying nonexistence of oscillatory mode. Condition (B9) is similar to the threshold pressure in Titze's wave model[19] in the sense that minimum energy (analogous to minimum lung pressure) is proportional to the fold damping coefficient and the prephonatory half-width.

[1] K. Ishizaka and J. L. Flanagan, ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' AT&T Tech. J. **51**, 1233–1268 (1972).
[2] I. Bogaert, ''Speech production by means of a hydrodynamic model and a discrete-time description,'' Inst Perception Res, Einhoven, the Netherlands, Report no. 1000, 1994.
[3] B. Story and I. Titze, ''Voice simulation with a body-cover model of the vocal folds,'' J. Acoust. Soc. Am. **97**, 1249–1260 (1995).
[4] I. Titze, ''The human vocal cords: A mathematical model, part I,'' Phonetica **28**, 129–170 (1973).
[5] I. Titze, ''The human vocal cords: A mathematical model, part II,'' Phonetica **29**, 1–21 (1974).
[6] F. Alipour and R. Scherer, ''Vocal fold bulging effects on phonation using a biophysical computer model,'' J. Voice **14**, 470–483 (2000).
[7] F. Alipour, D. Berry, and I. Titze, ''A finite-element model of vocal fold vibration,'' J. Acoust. Soc. Am. **108**, 3003–3012 (2000).
[8] X. Pelorson, A. Hirschberg, A. Wijnands, and H. Bailliet, ''Description of the flow through *in-vitro* models of the glottis during phonation,'' Acta Acust. (Beijing) **3**, 191–202 (1995).
[9] J. Flanagan, *Speech Analysis, Synthesis and Perception*, 2nd ed. (Springer-Verlag, New York, Berlin, 1972).
[10] L. Mongeau, N. Franchek, C. Coker, and R. Kubil, ''Characteristics of a pulsating jet through a small modulated oriface, with applications to voice production,'' J. Acoust. Soc. Am. **102**, 1121–1132 (1997).
[11] G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[12] T.-P. Liu, ''Transonic gas flow in a duct of varying area,'' Arch. Ration. Mech. Anal. **80**, 1–18 (1982).
[13] T.-P. Liu, ''Nonlinear stability and instability of transonic flows through a nozzle,'' Commun. Math. Phys. **83**, 243–260 (1982).
[14] R. Menikoff, K. Lackner, N. Johnson, S. Colgatem, J. Hyman, and G. Miranda, ''Shock wave driven by a phased implosion,'' Phys. Fluids A **3**, 201–218 (1991).
[15] I. Titze, ''Current topics in voice production mechanisms,'' Acta Oto-Laryngol. **113**, 421–427 (1993).
[16] J. Jiang and I. Titze, ''Measurement of vocal fold intraglottal pressure and impact stress,'' J. Voice **8**, 132–144 (1994).
[17] K. Stevens, *Acoustic Phonetics* (MIT, Cambridge, MA, 2000).
[18] R. Leveque, *Numerical Methods for Conservation Laws* (Birkhauser-Verlag, Basel, 1990).
[19] I. Titze, ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552 (1988).
[20] J. Golub and J. Ortega, *Scientific Computing and Differential Equations* (Academic, New York, 1992).
[21] R. Sataloff, ''The human voice,'' Sci. Am. **1992**, 108–115 (1992).
[22] S. Austin and I. Titze, ''The effect of subglottal resonance upon vocal fold vibration,'' J. Voice **11**, 391–402 (1997).
[23] G. Batchelor, *Introduction to Fluid Mechanics* (Cambridge U.P., Cambridge, MA, 1980).

Production benefits of childhood overhearing

Leah M. Knightly
Department of Psychology, University of California at Los Angeles, Los Angeles, California 90095

Sun-Ah Jun[a)]
Department of Linguistics, University of California at Los Angeles, Los Angeles, California 90095

Janet S. Oh
Department of Psychology, University of California at Los Angeles, Los Angeles, California 90095

Terry Kit-fong Au
Department of Psychology, University of Hong Kong, Hong Kong

(Received 15 August 2002; accepted for publication 7 April 2003)

The current study assessed whether overhearing Spanish during childhood helps later Spanish pronunciation in adulthood. Our preliminary report based on a subset of the data [Au *et al.*, Psychol. Sci. **13**, 238–243 (2002)] revealed that adults who overheard Spanish during childhood had better Spanish pronunciation, but not better morphosyntax, than adult learners of Spanish who had no childhood experience with Spanish. We now present data from the full sample with additional morphosyntax and pronunciation assessments, as well as measures to help rule out possible confounding prosodic factors such as speech rate, phrasing, and stress placement. Three groups of undergraduates were compared: 15 Spanish–English bilinguals (native Spanish speakers), 15 late learners of Spanish who overheard Spanish during childhood (childhood overhearers), 15 late learners of Spanish who had no regular experience with Spanish until middle or high school (typical late L2 learners). Results confirmed a pronunciation advantage for the childhood overhearers over the typical late L2 learners on all measures: phonetic analyses (VOT and degree of lenition), accent ratings (phoneme and story production), but no benefit in morphosyntax. Importantly, the pronunciation advantage did not seem attributable to prosodic factors. These findings illustrate the specificity of overhearers' advantage to phonological production. [DOI: 10.1121/1.1577560]

PACS numbers: 43.70.Ep, 43.70.Fq, 43.71.Hw [AL]

I. INTRODUCTION

Phonology is difficult for late second-language (L2) learners to master (Oyama, 1976; Tahta *et al.*, 1981; Flege and Fletcher, 1992; Flege *et al.*, 1995, 1999). Late learners' difficulty in producing nativelike accents in their L2 may in part be due to perceptual deficits (Flege, 1995). In fact, perceptual training in identification of L2 sounds seems to help late L2 learners pronounce the sounds of the target language better (Bradlow *et al.*, 1997, 1999).

Childhood exposure to the target language seems to benefit adult L2 perception and perhaps production as well. Case studies (Wode, 1981; Yamada, 1995) suggest that children learning English during visits to the U.S. maintained English-like production and perception after a two-year absence. Tees and Werker (1984) found that English-speaking adults who lived in a Hindi-speaking environment for the first year or two of their lives perceived Hindi phonemic contrasts reliably better than those who had no childhood experience with Hindi. However, it remains unclear what *type* of language experience (e.g., hearing, speaking) contributed to these perceptual and production benefits.

Infants learn much about the phonology of their ambient language simply by hearing it. They form language-specific vowel categories by 6 months of age and display ''language-specific phonetic'' perception by 10–12 months (Werker and Tees, 1984; Kuhl *et al.*, 1992; Werker, 1995). Impressively, 8-month-old infants learn word boundaries in a continuous speech stream after 2 min of exposure, using only statistical cues—where the only cue available for word boundaries is that sounds within a word are more likely to co-occur than sounds across word boundaries (Saffran *et al.*, 1996; Johnson and Jusczyk, 2001). Furthermore, such learning seems to be incidental rather than conscious and explicit. Saffran *et al.* (1997) asked adults and 6- and 7-year-old children to create pictures on a computer (cover task) while an artificial language played in the background. Even though participants were told they would not be tested on the acoustic material presented (i.e., the artificial language), both adults and children nonetheless demonstrated on a later test that they learned words of the artificial language after only 20 min of ''incidental'' exposure to it. Saffran *et al.* concluded that such ''incidental'' learning could be important in natural language acquisition.

Yet, to date few language acquisition studies have examined ''incidental'' learning *per se* outside the laboratory setting. Our study tries to fill this gap by focusing on ''incidental'' language learning in a *natural* language setting (i.e., overhearing everyday conversations). In an interim report of the current study, we explored possible effects of overhear-

[a)]Author to whom correspondence should be addressed. Electronic mail: jun@humnet.ucla.edu

ing Spanish during childhood (Au *et al.*, 2002). In that report, we compared three groups of English-speaking college students on their acquisition of Spanish phonology and morphosyntax. They included 10 Spanish-English bilinguals whose first language was Spanish (native speakers), 11 native English-speakers who regularly overheard Spanish during childhood but first learned Spanish in class around age 14 (childhood overhearers), and 12 native English-speakers who had minimal exposure to Spanish prior to learning it in class around age 14 (typical late L2 learners).

Compared to the typical late L2 learners, childhood overhearers' pronunciation of Spanish /p, t, k, b, d, g/ turned out to be more nativelike, as assessed by phonetic measures and accent ratings. Specifically, their VOT (voice-onset-time) for /p, t, k/ was reliably shorter for word initial position (e.g., *tacos*), thus closer to the norm of Spanish VOT, a short-lag VOT approximately 30–50 ms shorter than the long-lag VOT of English /p, t, k/ (Lisker and Abramson, 1964). Overhearers also produced Spanish intervocalic /b, d, g/, which are typically voiced and lenited (i.e., fricatives or approximants) but remain stops and are sometimes devoiced in English, more often as lenited consonants compared to the typical late learners. Importantly, overhearers' English pronunciation was not compromised; their pronunciation of English /p/ and /b/ was not reliably different from that of the typical late L2 learners.

Interestingly, the benefits of childhood overhearing did not extend to the area of morphosyntax. Au *et al.* found that childhood overhearers performed no better than typical late L2 learners in detecting morphosyntactic errors in a grammaticality judgment task or in their production of correct number and gender agreement in noun phrases. These interim findings suggest that simply hearing the language during childhood does not seem to benefit the mastery of morphosyntax, which like phonology is also challenging to late L2 learners (Snow and Hoefnagel-Hohle, 1978; Patkowski, 1980; Johnson and Newport, 1989; Johnson *et al.*, 1996; Flege *et al.*, 1999; Birdsong and Molis, 2001).

Although Au *et al.* (2002) found no benefits in morphosyntax, a more comprehensive assessment may yet show an overhearing advantage in this area. Additionally, the benefit in phonology reported by Au *et al.* could be due to confounding factors of prosody such as speech rate, placement of stress, and phrasing. Voiceless stop consonants tend to have longer VOT in slow speech, stressed syllables, and phrase initial position compared to those in fast speech, unstressed syllables, and phrase medial position, respectively (de Jong, 1995; Fougeron and Keating, 1997). Spanish intervocalic /b, d, g/ often become stops after a pause (Stockwell and Bowen, 1965), and slow speech typically contains more pauses than fast speech. In Au *et al.*'s study, participants were asked to produce word initial and medial /p, t, k, b, d, g/ in stressed syllables of Spanish words embedded in the carrier phrase *Diga* ___ *por favor* (meaning "Say ___ please"). Overhears could have produced more nativelike (i.e., shorter) VOT for word initial /p, t, k/ if the typical late learners happened to produce the carrier sentence more slowly than the overhearers, or if the overhearers misplaced stress on the Spanish words more often than the typical late learners (thereby producing more target phonemes without stress). Similarly, typical late learners' less frequent lenition of Spanish /b, d, g/ could have resulted from less fluent speech and hence more pauses before the target phonemes.

To reevaluate Au *et al.*'s (2002) interim findings, we report here the originally planned study in its entirety with the full sample, along with more in-depth phonetic analyses (degree of voicing in voiced stops, measurement of prosodic factors, as well as VOT and stop closure of voiceless stops, and lenition of voiced stops), additional accent rating data (narrative production), and additional assessment tasks in morphosyntax (verb morphology and story telling). In sum, the study presented here takes a more comprehensive look at whether there is a childhood overhearing advantage specific to phonology and not apparent in morphosyntax.

II. METHOD

A. Participants

Undergraduates enrolled in second year Spanish language courses at UCLA were recruited to complete a detailed language background questionnaire and follow-up interview about their experience with Spanish from birth to the time of testing. All participants completed a consent form and were paid for their participation. Out of the 238 participant responses received, 15 were identified as *childhood overhearers* and 100 were identified as *typical late L2 learners*. From the latter group, 15 were randomly selected with the constraint that they matched the *childhood overhearers* by gender and Spanish language instructor. Agreement between two trained coders on group assignment (childhood overhearer versus typical late L2 learner versus unclassifiable) was excellent (91% agreement; Cohen's Kappa$=0.90$). An additional 15 UCLA undergraduate native Spanish speakers were recruited for the *native speaker* group.[1] Three groups of 15 speakers (ten women, five men) made up the final sample.

Table I presents demographics and several language use measures for these three participant groups. Childhood overhearers and typical late L2 learners were all born in the U.S. and first learned Spanish in middle or high school. The overhearers reported overhearing Spanish during childhood from a parent/relative regularly and being spoken to in Spanish and speaking Spanish minimally (limited to short phrases and words in Spanish). Their amount of self-reported experience overhearing Spanish during ages 0–6 differed reliably from zero [$t(14)=2.33$, $p<0.05$] whereas their being addressed to in Spanish and speaking Spanish during ages 0–6 did not [t's$(14)<1.58$, n.s.]. Typical late L2 learners had minimal, if any, childhood exposure to Spanish. Native speakers were Spanish–English bilinguals who learned English as their L2 before age 10, and most of them were born in the U.S. Both native speakers and childhood overhearers were of Mexican or Central American descent. Overhearers were mostly mixed-Latino, with Spanish-English bilingual parents and relatives, whereas native speakers were mostly full-Latino, with monolingual Spanish-speaking parents and relatives. Typical late L2 learners were of non-Latino descent, with monolingual English-speaking parents and relatives.

TABLE I. Summary of speaker demographics and language use measures.

	Native speakers	Childhood overhearers	Typical late L2 learners
First language	Spanish	English	English
Age (years)	22.3[a] (0.7)	19.9[b] (0.6)	18.7[b] (0.3)
Hrs/wk heard Spanish			
age 0–6	33.0[a] (0.0)	5.3[b] (2.3)	0[c] (0)
age 6–12	30.1[a] (2.0)	1.8[b] (0.7)	0.13[b] (0.09)
Years taken Spanish classes			
middle/high school	2.4[a] (0.3)	3.9[b] (0.3)	4.4[b] (0.3)
college	0.9[a] (0.4)	1.0[a] (0.2)	0.6[a] (0.1)
No. visits to a Spanish-speaking country	6.7[a] (1.5)	1.5[b] (0.5)	0.1[b] (0.09)
Reported % use of Spanish (vs English)			
during high school years	41.7[a] (3.5)	9.3[b] (2.7)	9.9[b] (2.9)
during college years	30.4[a] (3.8)	7.5[b] (1.5)	8.0[b] (1.6)
Reported degree of (Phinney, 1992)			
identification with ethnic group	3.9[a] (0.04)	3.2[b] (0.1)	3.3[b] (0.2)
participation in Latino practices	3.5[a] (0.1)	2.9[b] (0.1)	2.4[c] (0.2)
Slang task performance (% correct)	80.9[a] (2.6)	19.8[b] (2.4)	2.7[c] (0.9)

Note. Table indicates group means with standard errors in parentheses. Means with different superscripts within a row were reliably different from each other according to Tukey's HSD test, $p<0.05$. Phinney Ethnic Identity Measures (1992) were on a four-point scale with higher numbers indicating more identification and participation.

Childhood overhearers and typical late L2 learners' self-reports were corroborated by reports from independent informants who knew the participants' prior experience with Spanish (e.g., parents). Reports on 40% of the research participants confirmed childhood overhearers' regular passive exposure to Spanish and limited spoken Spanish (single words, short phrases) and typical late L2 learners' lack of childhood exposure to Spanish. For further corroboration, participants' knowledge of Mexican/Central American household childhood expressions was tested to assess their childhood exposure to Spanish in the home. Participants read 20 English expressions (e.g., *cry baby, pacifier, dry crust in eyes*)—one at a time—on a computer screen and were asked to translate them into informal Spanish as they heard them at home, in the neighborhood, or in a schoolyard (slang production). They also heard 40 Spanish slang terms (e.g., *chiqueado* meaning ''spoiled child;'' *las escondidas* meaning ''hide-and-seek'') via a headset and were asked to translate them into English (slang comprehension). Participants' responses were audio-recorded and later independently transcribed and coded by two research assistants who were native Spanish speakers (average agreement between coders $=96\%$, disagreements were resolved by a third native Spanish speaker). As seen in Table I, the results corroborated participants' self-reports of childhood experience with Spanish quite well, suggesting that the childhood overhearers knew far less Spanish childhood slang than the native speakers, but nonetheless knew reliably more than the typical late L2 learners.

B. Phoneme production assessment

1. Stimuli and procedure

As in Au *et al.* (2002), pronunciation of Spanish phonemes /p, t, k, b, d, g/ was assessed. Twelve categories were created (3 places of articulation$\times$2 types of voicing$\times$2 positions in word) with three target words per category and each word containing the target phoneme in a stressed syllable (e.g., *base* ''base,'' *cabeza* ''head;'' see Table II for a complete list). Participants were asked to say each of the 36 target words in the sentence frame, ''*Diga* (target word) *por favor*,'' meaning ''Say (target word) please,'' thus producing 36 target sentences.

To see if overhearing Spanish during childhood might compromise speakers' English pronunciation, participants were asked to produce English voiceless consonant /p/ (in *pepper, pocket*) and voiced consonant /b/ (in *beggar, bonnet*) in the sentence frame, ''Take a (target word) once again.''

In order to lessen the potential prosodic confounds discussed earlier, participants were asked to stress the target word in each sentence and avoid pausing between words. Participants were given two practice sentences before reading the target sentences. Each sentence was presented three times in random order, and displayed on the computer screen for three seconds.

Instructions and stimuli for all tasks were presented on a Macintosh Powerbook G3 or 3400c/200 using PsyScope (Cohen *et al.*, 1993), with auditory stimuli presented via a headset, and participants responding via a button box. Participants were tested individually in a soundproof room, and their utterances were recorded using a Sennheiser HMD 25-1 microphone headset and a Marantz PMD-222 or PMD-430 professional recorder.

2. Measurement

Speech recordings were digitized at 12.5 kHz, and phonetic measurements were made using KAY Elemetrics speech analysis programs *CSL* and *MultiSpeech*. All measurers/coders (one primary, two secondary) were blind to the speakers' language backgrounds and analyzed roughly the same percentage of speakers from each of the three groups (native speakers, childhood overhearers, typical late

TABLE II. Spanish word list.

Target phoneme	Word initial	English gloss	Word medial	English gloss
/b/	base	base	cabeza	head
	beca	scholarship	jabón	soap
	beso	kiss	sabor	taste
/d/	datos	data	nadar	swim
	deja	to leave	pedido	an order
	día	day	rodar	roll
/g/	gallo	rooster	hogar	home
	gato	cat	pagó	paid
	goma	glue	pegó	hit
/p/	pase	to pass/pass	zapeta	diaper
	pena	embarrassed	vapor	vapor
	peso	weight	tapón	stopper
/t/	tacos	tacos	matar	to kill
	teja	shingling/weave	metido	it's in
	tía	aunt	notar	to notice
/k/	callo	a blister	tocar	to touch
	caso	case/pay attention	sacó	took out
	coma	command to eat/eat	pecó	sinned

L2 learners). To assess agreement, two speakers analyzed by each secondary measurer/coder were randomly selected to be analyzed by the primary measurer/coder.[2]

a. VOT and stop closure. For /p, t, k/, VOT was measured from the stop release to the onset of the second formant (F2) of the following vowel. Stop closure duration was measured from the offset of F2 of the vowel before the target phoneme to the release of the target phoneme. Mean differences between the primary and secondary measurer(s) for individual speakers were minimal for VOT (ranging from 2.6 to 3.6 ms for Spanish; 2.3 to 2.7 ms for English) and stop closure duration (ranging from 3.9 to 13.9 ms for Spanish; 2.1 to 2.3 ms for English).

b. Lenition. Spanish voiced consonants /b, d, g/ were categorized as either stops or lenited consonants, distinguished by an abrupt or a gradual change, respectively, in amplitude between the consonant and the following vowel. Percent agreement on these categorical assessments was high, ranging from 91% to 100% (Cohen's Kappas between 0.70 and 1.0).

c. Voicing. To assess the production of voiced consonants more comprehensively, we examined the degree of voicing during the consonant by categorizing all target consonants as having one of three voicing types: voiceless, partial voicing, or full voicing. For stops, voicing type was determined by voicing during the closure; for lenited consonants, by voicing during the duration of the consonant. Tokens were classified as ''voiceless'' if there was no voicing (i.e., no voice bar in spectrogram), as ''full voicing'' if voicing was present throughout the entire consonant duration, and as ''partial voicing'' if they displayed some voicing but not full voicing. Percent agreement on voicing among the coders was high for Spanish, ranging from 82% to 94% (Cohen's Kappas between 0.78 and 0.92) and for English, 83% to 92% (Cohen's Kappas 0.71 to 0.80).

d. Prosodic factors (Speech rate, phrasing, and stress placement). Speech rate was assessed by the duration from the F2 onset of the vowel after the release burst of the first word in the carrier sentence (e.g., /i/ in ''diga,'' /e/ in ''take a'') to the F2 offset of the vowel before the target word (e.g., /a/ in ''diga,'' /ə/ in ''take a''). Mean differences between measurers on individual speakers for speech rate were on average 9.4 ms for Spanish and 2.6 ms for English.

To measure phrasing differences, a coder listened to each Spanish token and judged whether a phrase boundary had been inserted before the target phoneme (subjective measure). As a quantitative measure a ratio of stop closure duration to speech rate duration was calculated for each token categorized as a stop (see Sec. II B 2 b). Since stop consonants in phrase initial position tend to have longer closure duration, most of the tokens judged to have a prosodic boundary had a ratio around 0.75, and those without around 0.50. Using this criterion, all tokens were reevaluated so that final classification was based on ratio (objective measure); only tokens with a ratio of 0.75 or greater were classified as having a phrase boundary before the target sound.

For stress misplacement, a coder listened to each Spanish token and noted whether speakers placed stress on the incorrect syllable (e.g., correct—*PAse*, incorrect—*paSE*).

C. Accent ratings

As in Au *et al.* (2002), we asked another group of native Spanish speakers to rate the participants' pronunciation of the target phonemes /p, t, k, b, d, g/ in the target sentences (phoneme accent ratings). To assess their accents in more natural speech, we asked yet another group of native Spanish speakers to rate participants' accents in narrative production (narrative accent ratings).[3]

a. Phonemic accent ratings. Forty-eight native Spanish speakers[4] were recruited to rate participants' pronunciation of the target phonemes in the 36 target sentences (e.g., ''*Diga*

pase por favor.''). The second of the three tokens produced by each speaker for each of the 36 target sentences was selected for rating. To avoid rater fatigue, 36 blocks were created whereby each block consisted of the same target sentence spoken by all speakers. Blocks were arranged into three sets so that each set included 12 blocks of target sentences containing the six target phonemes (/p, t, k, b, d, g/) in word-initial and medial position. In this way, three tokens of a category (e.g., tokens *pase*, *pena*, *peso* of word-initial /p/ category) were dispersed among the three sets (see ''stimuli and procedure,'' Sec. II B 1). Each rater was asked to rate each speakers' pronunciation of the target sound (e.g., the /p/ in *pase*)[5] for one of the three sets. Prior to rating, raters listened to each speaker say, ''*Diga teja por favor*,'' to familiarize themselves with the range of speakers' pronunciation abilities. A rating scale was presented on the computer screen during the familiarization and accent rating trials [1 = very strong foreign accent, definitely non-native; 2 = strong foreign accent; 3 = noticeable foreign accent; 4 = slight foreign accent; 5 = no foreign accent, definitely native; adopted from Bongaerts *et al*. (1997)]. Sentences within blocks were randomized and the test was self-paced. Interrater reliability was assessed using all of the ratings to compute, for each rater, an average rating for each speaker. An intraclass correlation on these averages revealed excellent agreement (average accent scores for individual participants: intraclass $R = 0.98$, $p < 0.0001$).

b. Narrative accent ratings. To assess participants' accents in more natural speech, we elicited narratives using a 10-page abridged version of a wordless children's picture book in Mercer Mayer's ''Frog, Where are you?'' series—a widely-used task for eliciting narratives from children and adults in different languages (e.g., Berman *et al*., 1994). Participants were given 2 min to scroll through the storybook pictures on a computer before coming up with a story. They then saw each page for 12 s, allowing time to say one or two sentences in Spanish per page. Audio recordings were independently rated by two native Spanish speakers[6] using the same scale adopted for phonemic accent ratings. The intraclass correlation between the two raters was excellent ($R = 0.91$).

D. Morphosyntax assessment

In addition to the grammaticality judgment task and noun-phrase production task reported in Au *et al*., the narrative production task just described and a verb-phrase production task were included to yield a more comprehensive picture of participants' mastery of Spanish grammar. For all tasks reported in this section, stimuli are available upon request.

a. Grammaticality judgment task. Participants listened to 33 grammatical–ungrammatical sentence pairs spoken by a native Spanish speaker. They heard the sentences in random order; they heard each sentence twice and pressed a button to indicate whether it was grammatical or ungrammatical. Both decision and response time were recorded. To minimize fatigue, the 66 sentences were presented in two blocks intermixed with other tasks. Ungrammatical sentences contained an error in one of the following categories: number/gender agreement in noun phrases (e.g.,**la flores*,**el carro blanca*), number agreement in verbs (e.g.,**Marta corren*,**mi mamá toman*), tense-aspect marking in verbs (e.g.,**Dentro de cuatro años*,**soy un abogado*), negation (e.g.,**El conoce a nadie*), indirect object (e.g.,**El enseña a nosotros*), or person agreement (e.g.,**nosotros comienzan*,**A qué hora llegué usted*).

b. Noun-phrase production task. Participants were asked to verbally complete five simple four-piece jigsaw puzzles designed to elicit four combinations of number and gender markers [adapted from Plann (1979)]. Each puzzle appeared on the computer screen for 18 s with four puzzle pieces and a puzzle frame (showing numbered spaces for the pieces). For example, pieces in one puzzle depicted two white pianos (*los pianos blancos*: plural masculine), two white cows (*las vacas blancas*: plural feminine), a black piano (*el piano negro*: singular masculine), and a black cow (*la vaca negra*: singular feminine). To complete the puzzle properly, participants had to specify the number and gender of the nouns used for naming the puzzle pieces (e.g., ''Pon los pianos blancos en cuatro, pon la vaca negra en tres,...'' meaning ''Put the white pianos in four, put the black cow in three,...''). Two native speakers of Spanish independently transcribed the audiotaped responses and coded them for number and gender agreement. A third native speaker resolved any discrepancies between the two transcribers/coders. Percent agreement between transcribers/coders was greater than 95%.

c. Verb-phrase production task. The task was adapted from Curtiss and Yamada's (1987) CYCLE test to elicit verb morphology (tense, aspect, person, and number) production. Participants heard 20 incomplete sentences, one at a time, illustrated with pictures presented on a computer. They were asked to offer sensible completions. For instance, they might hear ''Ayer fuí a la tienda, y yo...'' (meaning ''Yesterday I went to the store, and I...'') and see a picture of someone in a store. Next they would see a picture of the person buying milk. They were then given 6 s to complete the sentence. To be counted as an acceptable completion, appropriate morphosyntactic markers had to be used for the verb, as constrained by the lead-in clause (e.g., first person and singular form in the preterite tense in Spanish in the example just given). The 20 items were designed to elicit a variety of tense/aspect, number, and person markings. Agreement between the two transcribers/coders on participants' audiotaped responses was excellent (>90% agreement), and discrepancies were resolved by a third native speaker of Spanish.

d. Narrative production. Each frog story produced for the narrative accent rating task (see Sec. II C b) was rated by two native speakers of Spanish on a grammatical well-formedness scale, with 1 = definitely nonnative and 5 = definitely native. Interrater reliability was excellent (intraclass $R = 0.90$, $p < 0.0001$).

III. RESULTS

A. Phonology

1. Spanish phonemes

a. Prosodic factors. One-way ANOVAs revealed no reliable differences between the three speaker groups in any of

TABLE III. Prosodic factors results.

Measure	Native speakers	Childhood overhearers	Typical late L2 learners
Speech rate (ms)			
word initial	209 (5)	203 (7)	196 (7)
word medial	195 (6)	203 (7)	198 (8)
Misplaced stress (%)			
word initial	0.5 (0.4)	1.5 (0.5)	2.2 (1.1)
word medial	5.6 (2.0)	7.2 (1.9)	10.4 (2.5)
Phrase boundary before target word (%)			
word initial	16.2 (5.8)	25.2 (7.0)	15.7 (5.0)
word medial	7.8 (2.0)	8.6 (2.7)	7.2 (2.2)

Note. Standard errors are given in parentheses.

the prosodic factors examined [speech rate, percentage of misplaced stress, and percentage of prosodic boundary before the target word; F's$(2,42)<1.30$, n.s.; see Table III]. These findings suggest that any differences found among the three speaker groups in VOT, percent lenition, and voicing cannot be attributed to these prosodic factors. Rather, the phonetic values seem to reflect the degree of mastery of Spanish phoneme production.

b. Voiceless consonants. Childhood overhearers produced word initial /p, t, k/ with shorter VOT (i.e., more nativelike) than did the typical late learners. An ANOVA with place of articulation (i.e., bilabial/alveolar/velar) and consonant position (i.e., word initial/medial) as within-subject factors and speaker group as a between-subject factor on VOT revealed a main effect of speaker group [$F(2,42)=6.28$, $p<0.01$], place [$F(2,84)=149.91$, $p<0.001$], and a reliable position by group interaction [$F(2,42)=8.44$, $p<0.01$]. No other reliable main effect or interactions were found.

Figure 1 shows the mean VOT for word initial and medial /p, t, k/ for each speaker group. The typical late L2 learners produced reliably longer VOT in word initial position than did both the native speakers and overhearers (by HSDs, p's<0.01), but the three groups did not differ reliably in VOT for word medial /p, t, k/, thereby yielding the reliable group by position interaction. The main effect of place of articulation was due to longer VOT as the target sounds moved from a bilabial to velar position for all three groups (by Bonferroni, p's<0.001).

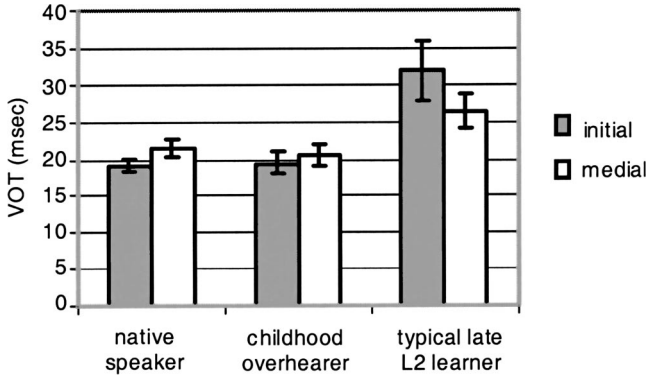


FIG. 1. Mean VOT duration of Spanish word initial and medial /p, t, k/ for each speaker group.

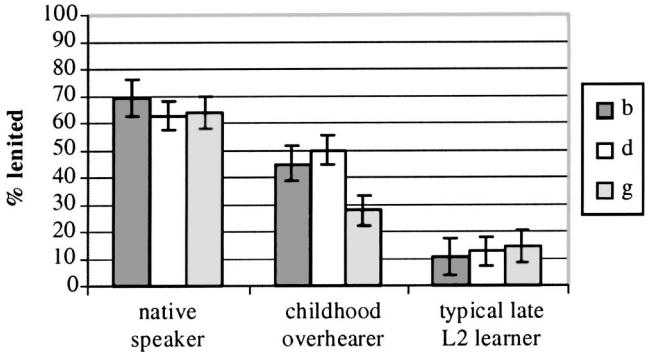


FIG. 2. Mean percentage use of lenited consonants in each place of articulation (i.e., /b/, /d/, /g/) for each speaker group.

One-way ANOVAs on speakers' stop closure duration for initial and medial /p, t, k/ revealed no reliable effects [F's$(2,42)<2.18$, n.s.].

c. Voiced consonants: degree of lenition. Speakers' tendency to produce /b, d, g/ as lenited consonants was computed based on the total number of phonetically voiced tokens for a particular phoneme (e.g., /b/, /d/, or /g/) and the number of tokens produced as lenited consonants. An ANOVA was then performed on these lenition percentages with place of articulation and position in word as within-subject factors, and speaker group as a between-subject factor. There were reliable main effects of speaker group [$F(2,42)=24.03$, $p<0.001$], position [$F(1,42)=74.78$, $p<0.001$], and place [$F(2,84)=3.76$, $p<0.05$], and a reliable interaction between place and group [$F(4,84)=5.65$, $p<0.01$].

Bonferroni *posthoc* tests showed that all speaker groups produced reliably more lenited consonants in word medial position ($M=52.9\%$, s.e.$=3.4\%$) than in word initial position ($M=26.4\%$, s.e.$=3.5\%$; $p<0.001$). As can be seen in Fig. 2, the place by group interaction was primarily due to overhearers being less likely to produce /g/ as a lenited consonant than /b/ and /d/ [t's$(14)=2.65$, p's<0.05]. This trend was not seen among the native speakers or the typical late L2 learners. Additional follow-up ANOVAs revealed that native speakers outperformed overhearers in producing intervocalic lenited /b/s, who in turn outperformed the typical late L2 learners [HSDs, p's<0.05; $F(2,42)=20.84$, $p<0.001$]. For intervocalic /d/, the native speakers and overhearers were comparable, and both outperformed the typical late L2 learners [HSDs, p's<0.001; $F(2,42)=21.91$, $p<0.001$]. In contrast, overhearers were no different from the typical late L2 learners in producing intervocalic lenited /g/s, and both did worse than the native speakers [by HSDs, p's<0.001; $F(2,42)=18.98$, $p<0.001$].

d. Voiced consonants: degree of voicing. Degree of voicing was assessed by the percentage of /b, d, g/ tokens produced by each speaker in each of the three voicing categories (i.e., voiceless, partial voicing, and full voicing), with the total number of /b, d, g/ tokens as the denominator. One-way ANOVAs revealed reliable group differences for partial [$F(2,42)=7.85$, $p<0.01$] and full voicing [$F(2,42)=6.40$, $p<0.01$] but not for voiceless. Tukey *posthoc* tests

TABLE IV. Accent rating assessment results.

Rated speech type	Native speakers	Childhood overhearers	Typical late L2 learners
Phonemic			
/p, t, k/	4.4[a] (0.08)	3.6[b] (0.08)	3.0[c] (0.09)
/b, d, g/	4.4[a] (0.10)	3.4[b] (0.10)	2.8[c] (0.10)
Narrative	5.0[a] (0.0)	3.0[b] (0.23)	2.4[c] (0.17)

Note. Accent ratings were made using a five-point scale, with higher ratings indicating more nativelike pronunciation. Standard errors are given in parentheses. Within a row, means with different superscripts were reliably different from each other according to Tukey's HSD test with $p<0.001$ for phonemic ratings, and with $p<0.05$ for narrative ratings.

revealed that native speakers produced /b, d, g/ with full voicing ($M=81.7\%$, s.d.$=15.2\%$) more often than typical late L2 learners ($M=55.5\%$, s.d.$=21.0\%$; by HSD, $p<0.01$). Overhearers' percent use of full voicing ($M=70.5\%$, s.d.$=23.4\%$) was in between these two groups, although not reliably different from either. Native speakers used partial voicing less often than typical late L2 learners ($M=12.0\%$, s.d.$=3.0\%$; $M=29.5\%$, s.d.$=3.7\%$, respectively; by HSD, $p<0.01$). Again, overhearers were in between these two groups but differed from neither reliably ($M=13.7\%$, s.d.$=3.6\%$).

2. English phonemes

English production data for word-initial /p/ and /b/ were collected from 14 native speakers, 12 childhood overhearers, and 14 typical late L2 learners.[7] One-way ANOVAs revealed no reliable differences among the three groups in speech rate, closure duration, or VOT. All groups produced English /p/ as an aspirated stop with mean VOT values ranging from 50 to 60 ms and English /b/ at 10 ms. The three groups of speakers were also similar in their voicing of English /b/, producing it as a voiceless stop approximately 25% of the time.

B. Accent ratings

1. Phonemic accent ratings

Given the excellent interrater agreement (see Sec. II C a), a rating averaged across raters for each speaker was calculated for both voiced and voiceless consonants in word initial and medial position. ANOVAs on these average ratings with group as a between-subject factor and position as a within-subject factor revealed a reliable effect of group for both voiceless [$F(2,41)=71.76$, $p<0.001$] and voiced consonants [$F(2,41)=72.32$, $p<0.001$]. No other reliable main effect or interaction was found. Posthoc tests revealed that native speakers received reliably better accent ratings than overhearers, who in turn received reliably better ratings than typical late L2 learners [by HSDs, p's<0.001; see Table IV]. Pearson correlations between phonemic accent ratings and phonetic measurements, namely VOT and lenition, were substantial when averaged across tokens (VOT, $r=-0.53$, $p<0.001$; lenition, $r=0.78$, $p<0.001$) as well as at the level of individual tokens (i.e., second token; VOT, $r=-0.50$, $p<0.01$; lenition, $r=0.77$, $p<0.001$).

TABLE V. Participants' performance on Spanish morphosyntax assessment tasks.

Measure	Native speakers	Childhood overhearers	Typical late L2 learners
Grammaticality judgment			
percentage correct	91.8[a] (0.97)	63.6[b] (1.6)	62.5[b] (2.6)
reaction time (ms)	1201[a] (124)	2661[b] (283)	1936[a] (307)
Noun-phrase production			
gender agreement	94.8[a] (2.6)	66.3[b] (6.2)	72.7[b] (2.8)
number agreement	93.8[a] (2.6)	82.5[a] (5.3)	92.3[a] (1.7)
Verb-phrase production			
tense/aspect	94.1[a] (1.6)	50.4[b] (5.1)	50.0[b] (5.1)
person	98.2[a] (1.0)	68.6[b] (4.4)	72.3[b] (5.1)
number	98.6[a] (1.0)	70.7[b] (4.8)	80.3[b] (4.9)
Narrative production	4.93[a] (6.7)	2.5[b] (0.17)	2.6[b] (0.14)

Note. Narrative production ratings were on a 5-point scale, with high ratings indicating better formulated sentences. For all other measures, the table indicates the mean percentage correct unless otherwise specified. Standard errors are given in parentheses. Within a row, means with different superscripts were reliably different from each other according to Tukey's HSD test, $p<0.01$. Numbers with the same superscript were not reliably different from each other.

2. Narrative accent ratings

As shown in Table IV, similar results were found for the frog story accent ratings.[8] An ANOVA on the ratings averaged across raters for each speaker revealed reliable group differences [$F(2,41)=73.57$, $p<0.001$]. Native speakers were rated as having a better whole-sentence accent than the overhearers, who in turn were rated more favorably than the typical late L2 learners (by Tukey's HSDs, p's<0.05).

C. Morphosyntax

1. Grammaticality judgment

There were reliable group differences for grammaticality judgment both in terms of accuracy and reaction time [$F(2,41)=77.41$, $p<0.001$ and $F(2,41)=8.01$, $p<0.01$, respectively; see Table V]. Native speakers outperformed both childhood overhearers and typical late L2 learners in making correct grammaticality judgments (Tukey's HSDs, p's <0.01). However, the latter two did not differ reliably from each other. The overhearers, but not the typical late L2 learners, took longer than the native speakers in deciding whether a sentence was grammatical (Tukey's HSD, $p<0.01$) whereas the latter two did not differ reliably from each other.

2. Noun-phrase production

The native speakers outperformed both the overhearers and the typical late L2 learners in marking gender agreement, but not in marking number agreement [$F(2,35)=14.38$, $p<0.001$ and $F(2,35)=3.07$, $p=0.06$, respectively; see Table V for Tukey's HSD test results]. The childhood overhearers and the typical late L2 learners did not differ reliably from each other.

3. Verb-phrase production

For verb morphology production, the native speakers outperformed the childhood overhearers and the typical late

L2 learners, who again did not differ reliably from each other [$F(2,37)=9.6$, $p<0.001$; see Table V for Tukey's HSD test results].

4. Narrative production

Native speakers' frog stories were rated as containing fewer morphosyntactic errors than the overhearers' and the typical late L2 learners' stories. The latter two did not differ reliably from each other [$F(2,41)=118.27$, $p<0.001$; see Table V for Tukey's HSD test results].

Taken together, the results of this study reveal considerable benefits of childhood overhearing in phonology and no measurable benefits in morphosyntax.

IV. DISCUSSION AND CONCLUSIONS

Childhood overhearers of Spanish were found to have more nativelike Spanish pronunciation than typical late L2 learners according to phonetic analyses (i.e., VOT, lenition), phonemic accent ratings, and narrative accent ratings. Importantly, the overhearers' advantage in Spanish pronunciation was not found to be attributable to prosodic factors such as speech rate, phrasing of the carrier sentence, and/or the location of stress on the target word. Native speakers, childhood overhearers, and typical late L2 learners were all comparable on these prosodic characteristics. That said, these findings do not necessarily mean that the three groups acquired prosodic features of Spanish to the same degree. Studies on prosodic transfer (e.g., Ueyama and Jun, 1998; Jun and Oh, 2000) show that the degree of proficiency in an L2 is reflected in the realization of L2 intonation patterns and prosodic grouping of words. No reliable group differences were uncovered in the present study perhaps because the sentences evaluated for prosodic factors (i.e., the target sentences) were short and simple. The groups might have shown prosodic differences if we had evaluated production in the story-telling task, which elicited longer and more natural sentences with more complex syntactic structures.

Our study speaks to the relation between perception and production. Flege's speech learning model (Flege, 1995) postulates that a speaker's pronunciation ability should be only as good as his or her perceptual abilities. When compared to the typical late L2 learners, the childhood overhearers' better accents may reflect more nativelike perceptual discrimination of Spanish phonemes. Although our study did not test this hypothesis directly, a similar study on Korean L2 acquisition (Oh *et al.*, 2003) suggests that this may indeed be the case. Oh *et al.* found that adults who regularly heard Korean during childhood performed better in their perception of Korean stops than those who had no exposure to Korean until college. Childhood overhearing, then, may have helped improve speech perception. Such perceptual abilities can be put to use when childhood overhearers try to learn the overheard language later in life (e.g., in high school and college), resulting in a better accent. If this account is on the right track, our findings of childhood overhearing benefits in speech production can be added to the growing body of evidence for a perception-production link. That is, early overhearing triggers better perception, which can be translated into better production even for late language learners (e.g., Bradlow *et al.*, 1997, 1999).

However, the logical inference of "if better perception, then better production" needs elaboration. Note that the overhearers in this study, like the typical late L2 learners, rarely produced intervocalic /g/ as a lenited consonant. Nonetheless, they were nativelike for intervocalic /d/, and their intervocalic /b/ fell between these two levels of performance. This pattern of results may be due to the frequency with which the overhearers were exposed to lenited /b, d, g/. For example, lenited /g/ (i.e., [ɣ]) is not in the English sound system either phonemically or allophonically, so lack of practice may account for the overhearers' difficulty in producing lenited /g/. By contrast, the overhearers were nativelike for lenited /d/ (i.e., [ð]), which happens to be a high-frequency phoneme in English (e.g., "th" in *this, other*). For lenited /b/ (i.e., [β]), which is not an English phoneme but is close in articulation and similar in voicing to the English /v/, the overhearers produced it better than the typical late learners but not yet nativelike.

The puzzle then is why typical late L2 learners had great difficulty producing lenited sounds even when one of them occurs in English (namely, [ð]).[9] It may have to do with Spanish and English orthography. Since the letters "b, d, g" are produced as stop consonants in English, the typical late L2 learners could be more vulnerable to being misled by the orthographic presentation of Spanish /b, d, g/ than the childhood overhearers. Another possibility is that the typical late L2 learners had not yet acquired the allophones of Spanish /d/ in intervocalic position. That is, even though they knew how to produce the dental fricative, i.e., a lenited /d/, they were not aware that the same sound was also an allophone of /d/ in Spanish. Note that typical late L2 learners also produced Spanish voiced stops with less voicing than did native speakers. Childhood overhearers' voicing was in between these two groups although not reliably different from either. Combined, these findings hint at the possibility that childhood overhearing of Spanish may lead to better awareness of Spanish allophones.

Our findings are also relevant to the nature of input in early childhood bilinguals' L2 phonology. As in Au *et al.* (2002), we found that overhearers' pronunciation advantage did not come at the cost of producing English /p/ and /b/ with a Spanish accent. Nor did the English of our native Spanish speakers (i.e., Spanish–English bilinguals) seem compromised since they also produced English /p/ and /b/ with values comparable to that of typical late L2 learners. These findings are consistent with prior findings where Spanish–English bilinguals who learned English by age 5–6 produced English /t/ (Flege, 1991), /p/ (Williams, 1980), and /b/ (Flege and Eefting, 1988) like English monolinguals. However, our results contrast with findings where Spanish–English bilinguals produced English /p, t, k/ with "compromised" VOT values (Flege and Eefting, 1987) and English /b/ with Spanish-like voicing lead (Flege and Eefting, 1988).

One way to make sense of these conflicting results is to focus on the nature of English input. Flege and Eefting (1987) speculated that the early childhood bilinguals in their

study had ''compromised'' VOT values because much of their English input was Spanish-accented; they were born in the U.S. but were attending a university in a Spanish-speaking country at the time of testing. In contrast, the bilinguals in Williams (1980) and Flege (1991) were more like the Spanish–English bilinguals in our study in that they were living in the U.S. at the time of testing and used *both* languages in their daily lives. Thus, the nature (hearing good models of American English) as well as the timing of L2 input may play an important role in childhood bilinguals' eventual L2 pronunciation.

An important conclusion of the current study is that the childhood overhearing advantage seems to be domain specific: it is very robust in phonology but so far not detectable in morphosyntax—a domain, like phonology, that is easy for children to acquire but difficult for adults to master. Even with our rather comprehensive assessment of morphosyntax with additional tasks not included in Au *et al.*'s (2002) study (i.e., elicited verb-phrase production, story production), no childhood overhearing advantage in morphosyntax was detected. Perhaps it takes more than merely overhearing a language during childhood to gain an edge in morphosyntax acquisition. Future research can help determine whether other kinds of childhood language experience, such as speaking a language for a few years during childhood, results in measurable benefits in mastery of morphosyntax as well as phonology for adults (re-)learning a childhood language.

Our results can have important applied implications for immigrant children in a predominantly monolingual nation such as the U.S. Most immigrant children in such linguistic environments tend to lose, or severely limit their use of, their heritage language (e.g., Fillmore, 1991; Au, in press). Our study can help immigrant parents and early childhood educators make more informed decisions about what kind of childhood language environment and experience they want to foster for their children. Specifically, this study suggests that overhearing a language during childhood can probably translate into a measurable accent advantage when children try to acquire the overheard language later in life.

ACKNOWLEDGMENTS

We thank Celia Adame, Marilyn Alvarado, Brenda Alvarez, Argelia Andrade, Cecilia Arriaza, J. Patrick Barjam, Jessica Brauner, Mauricio Carvallo, Ivonne Cisneros, Susan Curtiss, Kristina Cutura, Olga Escamilla, Laura España, Catherine Fountain, Ingrid Gooding, John Grinstead, Vilma Hernandez, Gerardo Ibarra, Patricia Keating, Sahyang Kim, Hyuck-Joon Lee, Reuben Lim, Ann Marroquin, Olivia Martínez, Karla Mayen, Nancy Miranda, Claudia Parodi, Jaime Paz, Katie Polsky, Irasema Ramos, Salome Rincón, Krista Rodriguez-Bruno, Laura Romo, Norma Salazar, Susan Schaffer, Sylvia Sherno, Minjung Son, Henry Tehrani, Motoko Ueyama, and John Whalen for their help with various aspects of this research. We thank the research participants for their time. We also thank Anders Löfqvist and two anonymous reviewers for their insightful and valuable comments on an earlier version of this paper. This work was supported by NIMH Grant No. MH56118 (Au and Jun), an NIMH supplement (Au and Knightly), an NSF predoctoral fellowship (Oh), NIMH traineeships in Developmental Cognitive Science 5T32MH10926 (Knightly, Oh), and seed grants from University of California Linguistic Minority Research Institute, UCLA Institute of American Cultures, and UCLA Academic Senate (Au).

[1]Among the 45 speakers participating in the current study, 18 speakers (5 of the 15 native speakers, 6 of the 15 childhood overhearers, and 7 of the 15 typical late L2 learners) contributed data to Au *et al.*'s (2002) interim report, which also included 15 other speakers from a pilot study. (i.e., 5 natives, 5 childhood overhearers, and 5 typical late L2 learners).

[2]To assess inter-measurer/coder agreement, the primary measurer/coder analyzed the Spanish data on two speakers measured by each of the two secondary measurers/coders, and the English data on two speakers measured by the only secondary measurer/coder. Thus, the reported ranges of agreement are based on the four values for Spanish and two values for English.

[3]Except where noted, all native Spanish-speaking raters were of Mexican or Central American descent. In addition, speech samples assessed by two native Spanish speakers (Mexican) verified self-reports of raters' native speaker status.

[4]Of the 48 raters, 36 were of Mexican or Central American descent, 5 were from Spain, and 7 were of South American descent; 3 from Argentina and 1 each from Columbia, Peru, Ecuador and Bolivia.

[5]Even though raters were instructed to focus on the target segment, their ratings could be influenced by the carrier sentence. However, since the carrier sentence was short and repeated for every target sentence, its effect may be minimal. Compared to narrative accent ratings based on more varied and longer sentences, the phonemic accent ratings are likely to reflect the quality of the target phonemes.

[6]Since the narrative accent ratings involved a much smaller data set (45 story sessions) compared to the phonemic accent ratings, only two raters of Mexican American descent, who readily achieved excellent interrater reliability, were recruited for this task.

[7]One native speaker, three overhearers, and one typical late L2 learner were not available for testing when English data were collected about two months after the Spanish data collection.

[8]The relatively low phonemic accent ratings for native speakers may reflect the difficulty in rating just the target sounds embedded in the carrier sentence. The raters may therefore have been more conservative, using a narrower range of the rating scale and staying closer to the mid-point. On the other hand, raters may have been more confident about their narrative accent ratings based on the longer speech samples (i.e., stories) and hence showed more range in their ratings.

[9]Examination of individual lenition percentages revealed that two typical late L2 learners failed to produce any lenited consonants.

Au, T. K. (**in press**). ''Salving heritage languages,'' in *Heritage Language Acquisition: A New Field Emerging*, edited by D. Brinton and O. Kagan (Erlbaum, Mahwah, NJ).

Au, T. K., Knightly, L. M., Jun, S., and Oh, J. S. (**2002**). ''Overhearing a language during childhood,'' Psychol. Sci. **13**, 238–243.

Berman, R., Slobin, D. I., Aksu-Koc, A. A., Bamberg, M., Dasinger, L., Marchman, V., Neeman, Y., Rodkin, P. C., and Sebastian, E. (**1994**). *Relating Events in Narrative: A Cross Linguistic Developmental Study* (Erlbaum, Hillsdale, NJ).

Birdsong, D., and Molis, M. (**2001**). ''On the evidence for maturational constraints in second-language acquisition,'' J. Memory Lang. **44**, 235–249.

Bongaerts, T., Van Summeren, C., Planken, B., and Schils, E. (**1997**). ''Age and ultimate attainment in the pronunciation of a foreign language,'' Studies in Second Lang. Acquis. **19**, 447–465.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (**1999**). ''Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production,'' Percept. Psychophys. **61**, 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). ''Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (**1993**). ''PsyScope: A new graphic interactive environment for designing psychology

experiments,'' Behav. Res. Methods Instrum. Comput. **25**, 257–271.
Curtiss, S., and Yamada, J. (**1987**). ''The Curtiss-Yamada Comprehensive Language Evaluation (CYCLE),'' University of California, Los Angeles.
de Jong, K. (**1995**). ''The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation,'' J. Acoust. Soc. Am. **97**, 491–504.
Fillmore, L. W. (**1991**). ''When learning a second language means losing the first,'' Early Childhood Res. Qu. **6**, 323–346.
Flege, J. E. (**1991**). ''Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language,'' J. Acoust. Soc. Am. **89**, 395–411.
Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Perception*, edited by W. Strange (York, Baltimore, MD), pp. 233–272.
Flege, J. E., and Eefting, W. (**1987**). ''Production and perception of English stops by native Spanish speakers,'' J. Phonetics **15**, 67–83.
Flege, J. E., and Eefting, W. (**1988**). ''Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation,'' J. Acoust. Soc. Am. **83**, 729–740.
Flege, J. E., and Fletcher, K. L. (**1992**). ''Talker and listener effects on degree of perceived foreign accent,'' J. Acoust. Soc. Am. **91**, 370–389.
Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995**). ''Factors affecting strength of perceived foreign accent in a second language,'' J. Acoust. Soc. Am. **97**, 3125–3134.
Flege, J. E., Yeni-Komshian, G. H., and Liu, S. (**1999**). ''Age constraints on second-language acquisition,'' J. Memory Lang. **41**, 78–104.
Fougeron, C., and Keating, P. A. (**1997**). ''Articulatory strengthening at edges of prosodic domains,'' J. Acoust. Soc. Am. **101**, 3728–3740.
Johnson, E. K., and Jusczyk, P. W. (**2001**). ''Word segmentation by 8-month-olds: When speech cues count more than statistics,'' J. Memory Lang. **44**, 548–567.
Johnson, J. S., and Newport, E. L. (**1989**). ''Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language,'' Cognit Psychol. **21**, 60–99.
Johnson, J. S., Shenkman, K. D., Newport, E. L., and Medin, D. L. (**1996**). ''Indeterminacy in the grammar of adult language learners,'' J. Memory Lang. **35**, 335–352.
Jun, S.-A., and Oh, M. (**2000**). ''Acquisition of 2nd Language Intonation,'' in *Proceedings of International Conference on Spoken Language Processing* (Beijing, China), Vol. 4, pp. 76–79.
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 (Washington, DC, U.S.) **255**, 606–608.
Lisker, L., and Abramson, A. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.
Oh, J. S., Jun, S., Knightly, L. M., and Au, T. K. (**2003**). ''Holding on to childhood language memory,'' Cognition **86**, B53–B64.
Oyama, S. (**1976**). ''A sensitive period for the acquisition of a nonnative phonological system,'' J. Psycholinguist. Res. **5**, 261–283.
Patkowski, M. (**1980**). ''The sensitive period for the acquisition of syntax in second language,'' Lang. Learning **30**, 449–472.
Phinney, J. S. (**1992**). ''The multigroup ethnic identity measure: a new scale for use with diverse groups,'' J. Adolescent Res. **7**, 156–176.
Plann, S. J. (**1979**). ''Morphological problems in the acquisition of Spanish in an immersion classroom,'' in *The Acquisition and Use of Spanish and English as First and Second Languages*, edited by R. W. Anderson (TESOL, Washington, DC), pp. 119–132.
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). ''Statistical learning by 8-month-old infants,'' Science (Washington, DC, U.S.) **274**, 1926–1928.
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (**1997**). ''Incidental language learning: Listening (and Learning) out of the corner of your ear,'' Psychol. Sci. **8**, 101–105.
Snow, C. E., and Hoefnagel-Hohle, M. (**1978**). ''The critical period for language acquisition: evidence from second language learning,'' Child Dev. **49**, 1114–1128.
Stockwell, R. P., and Bowen, J. D. (**1965**). *The Sounds of English and Spanish* (Univ. of Chicago, Chicago).
Tahta, S., Wood, M., and Loewenthal, K. (**1981**). ''Foreign accents: Factors relating to transfer of accent from the first language to a second language,'' Lang. Speech **24**, 265–272.
Tees, R. C., and Werker, J. F. (**1984**). ''Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds,'' Can. J. Psychol. **38**, 579–590.
Ueyama, M., and Jun, S.-A. (**1998**). ''Focus realization in Japanese English and Korean English intonation,'' in *Japanese and Korean Linguistics* (CSLI, Cambridge U.P., Cambridge), Vol. 7, pp. 629–645.
Werker, J. F. (**1995**). ''Age-related changes in cross-language speech perception: Standing at the crossroads,'' in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Perception*, edited by W. Strange (York, Baltimore, MD), pp. 155–169.
Werker, J. F., and Tees, R. C. (**1984**). ''Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,'' Infant Behav. Dev. **7**, 49–63.
Williams, L. (**1980**). ''Phonetic variation as a function of second-language learning,'' in *Child Phonology Volume 2: Perception*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), pp. 185–215.
Wode, H. (**1981**). *Learning a Second Language: An Integrated View of Language Acquisition* (Gunter Narr Verlag, Tubingen).
Yamada, R. A. (**1995**). ''Age and acquisition of second language speech sounds: perception of American English /r/ and /l/ by native speakers of Japanese,'' in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Baltimore), pp. 305–320.

Dichotic speech recognition in noise using reduced spectral cues

Philipos C. Loizou[a] and Arunvijay Mani
Department of Electrical Engineering, University of Texas at Dallas, P.O. Box 830688, EC 33, Richardson, Texas 75083-0688

Michael F. Dorman
Arizona Biomedical Institute, Arizona State University, Tempe, Arizona 85287

(Received 30 August 2002; revised 25 April 2003; accepted 25 April 2003)

It is generally accepted that the fusion of two speech signals presented dichotically is affected by the relative onset time. This study investigated the hypothesis that spectral resolution might be an additional factor influencing spectral fusion when the spectral information is split and presented dichotically to the two ears. To produce speech with varying degrees of spectral resolution, speech materials embedded in +5 dB S/N speech-shaped noise were processed through 6–12 channels and synthesized as a sum of sine waves. Two different methods of splitting the spectral information were investigated. In the first method, the odd-index channels were presented to one ear and the even-index channels to the other ear. In the second method the lower frequency channels were presented to one ear and the high-frequency channels to the other ear. Results indicated that spectral resolution did affect spectral fusion, and the effect differed across speech materials, with the sentences being affected the most. Sentences, processed through six or eight channels and presented dichotically in the low–high frequency condition were not fused as accurately as when presented monaurally. Sentences presented dichotically in the odd–even frequency condition were identified more accurately than when presented in the low–high condition. [DOI: 10.1121/1.1582861]

PACS numbers: 43.71.Es, 43.71.Ky [KWG]

I. INTRODUCTION

Since the seminal paper by Cherry in the early 1950s (Cherry, 1953) on the ''cocktail party'' effect and dichotic listening, much work has been done to understand dichotic speech perception (Cutting, 1976). Several researchers have demonstrated the remarkable ability of the brain to fuse two different signals presented simultaneously in both ears (i.e., presented dichotically) to a single auditory percept. Fusion of speech signals can occur in many forms (Cutting, 1976). Broadbent and Ladefoged (1957) demonstrated that when the listeners were presented simultaneously with a signal containing the F1 of /da/ to the left ear and a signal containing the F2 of /da/ to the right ear, subjects heard the stimulus as /da/. Others (e.g., Hawles, 1970; Cutting, 1976) have demonstrated that if a particular /ba/ is presented to one ear and a particular /ga/ is presented to the other ear, listeners often reported hearing a single item, /da/. Sound localization is yet another form of fusion since the acoustic signals received in each ear differ in amplitude and phase (time difference). Cutting (1976) identified and analyzed six different types of fusion that occur at possibly three different levels of auditory processing. Cutting termed the fusion observed in the Broadbent and Ladefoged (1957) study as ''spectral'' fusion, and that is the focus of this paper. Spectral fusion occurs when different, but complementary, spectral regions of the same signal are presented to opposite ears.

Much work has been done to understand the various factors influencing spectral fusion (Cutting, 1976). Three factors were found to be most important: relative onset time, relative intensity, and fundamental frequency (F0). Cutting (1976) delayed the information reaching one ear with respect to the other, and found that spectral fusion decreased significantly when the delay was increased more than 40 ms. Rand (1974) found no effect on fusion when the signal presented to one ear was attenuated by as much as 40 dB, suggesting that spectral fusion is immune to relative intensity variations. Similar experiments were carried out to examine the effect of differences in fundamental frequency (F0) on spectral fusion. When listeners were presented with signals differing in F0, they reported hearing two sounds, however the identity of the fused percept was maintained independent of differences in F0 (Darwin, 1981; Cutting, 1976). Darwin (1981) demonstrated that the identification of ten three-formant vowels was unaffected by formants excited at different fundamentals or starting at 100-ms intervals.

In summary, of the three factors investigated in the literature, the relative onset-timing difference seemed to have the largest effect on dichotic speech perception. In this paper, we investigate whether spectral resolution could be considered as another factor that could potentially influence spectral fusion. Spectral resolution is an issue that needs to be taken into account when dealing with cochlear implant listeners, known to receive a limited amount of spectral information. The recent introduction of bilateral cochlear implants spurred the question of whether cochlear implant listeners would be able to fuse speech information presented dichotically. The possible advantage of dichotic (electrical) stimu-

[a]Electronic mail: loizou@utdallas.edu

TABLE I. The 3-dB cutoff frequencies (Hz) of the bandpass filters used in this study. F_L and F_H indicate the low and high cutoff frequencies respectively of the bandpass filters.

	6 channels		8 channels		12 channels	
Channels	F_L	F_H	F_L	F_H	F_L	F_H
1	300.0	487.2	261.9	526.6	191.6	356.8
2	487.2	791.1	526.6	857.3	356.8	549.5
3	791.1	1284.5	857.3	1270.4	549.5	774.3
4	1284.5	2085.8	1270.4	1786.5	774.3	1036.6
5	2085.8	3387.1	1786.5	2431.2	1036.6	1342.5
6	3387.1	5500.0	2431.2	3236.7	1342.5	1699.4
7			3236.7	4242.9	1699.4	2115.8
8			4242.9	5500.0	2115.8	2601.5
9					2601.5	3168.2
10					3168.2	3829.2
11					3829.2	4600.4
12					4600.4	5500.0

lation is reduction in channel interaction, since one can stimulate the electrodes alternately across the two ears (e.g., electrode 1 in left ear followed by electrode 2 in right ear and so on).

To examine the effect of spectral resolution on dichotic speech recognition, noisy speech was processed into a small number (6–12) of channels and presented to normal-hearing listeners dichotically. Two different conditions were considered. In the first condition, low-frequency information was presented to one ear and high-frequency information was presented to the other ear. In the second condition, the frequency information was interleaved between the two ears with the odd-index frequency channels fed to one ear and the even-index channels fed to the other ear. At issue is whether spectral fusion is affected by (1) poor spectral resolution or/and (2) the way spectral information is split (low/high versus interleaved) and presented to the two ears. We hypothesize that both spectral resolution and the type of spectral information presented to the two ears will affect spectral fusion.

II. DICHOTIC LISTENING IN NOISE

A. Method

1. Subjects

Nine normal-hearing listeners (20 to 30 years of age) participated in this experiment. All subjects were native speakers of American English. The subjects were paid for their participation. The subjects were undergraduate students from the University of Texas at Dallas.

2. Speech material

Subjects were tested on sentence, vowel, and consonant recognition. The vowel test included the syllables: ''heed, hid, hayed, head, had, hod, hud, hood, hoed, who'd, heard'' produced by male and female talkers. A total of 22 vowel tokens were used for testing, 11 produced by seven male speakers and 11 produced by six female speakers (not all speakers produced all 11 vowels). These tokens were a subset of the vowels used in Loizou *et al.* (1998) and were selected from a large vowel database (Hillenbrand *et al.*, 1995) to represent the complete area of the vowel space. The consonant test used 16 consonants in /aCa/ context taken from the Iowa consonant test (Tyler *et al.*, 1987). All /aCa/ syllables were produced by a male speaker.

The sentence test used sentences from the HINT database (Nilsson *et al.*, 1994). Two lists, consisting of 20 sentences, were used for each condition, and different lists were used in all conditions.

3. Signal processing

Speech material was first low-pass filtered using a sixth-order elliptical filter with a cutoff frequency of 6000 Hz. Filtered speech was passed though a preemphasis filter (high-pass) with a cutoff frequency of 2000 Hz. This was followed by band-pass filtering into n ($n=6,8,12$) frequency bands using sixth-order Butterworth filters. The cutoff frequencies of the bandpass filters are given in Table I. Logarithmic frequency spacing was used for $n=6$, and mel frequency spacing (linear spacing up to 1000 Hz and logarithmic thereafter) was used for $n=8,12$. The output of each channel was passed through a full-wave rectifier followed by a second-order Butterworth low-pass filter with a cutoff frequency of 400 Hz to obtain the envelope of each channel output. Corresponding to each channel a sinusoid was generated with frequency set to the center frequency of the channel and with amplitude set to the root-mean-squared (rms) energy of the channel envelope, estimated every 4 ms. The phases of the sinusoids were estimated from the FFT of the speech segment as per Loizou *et al.* (1999). No interpolation was done on the amplitudes or phases to smooth out any discontinuities across the 4-ms segments.

Two sets of sine waves were synthesized for dichotic presentation, one corresponding to the left ear and one corresponding to the right ear. The sine waves with frequencies corresponding to the left-ear channels were summed to produce the left-ear signal, and, similarly, the sinewaves corresponding to the right-ear channels were summed to produce the right-ear signal. In the condition, for instance, in which the low-frequency information was presented to the left ear and the high-frequency information was presented to the right ear, the envelope amplitudes corresponding to channels $1-n/2$ were used to synthesize the left-ear signal and the amplitudes corresponding to channels $n/2+1-n$ were used

to synthesize the right-ear signal. The levels of the two synthesized signals (left and right) were adjusted so that the sum of the two levels was equal to the rms value of the original speech segment. This was done by multiplying the left and right signals by the same energy normalization value. Hence, no imbalance was introduced, in terms of level differences or spectral tilt, between the left and right envelope amplitudes.

4. Procedure

The experiments were performed on a PC equipped with a Creative Labs SoundBlaster soundcard. Stimuli were played to the listeners either monaurally or dichotically through Sennheiser's HD 250 Linear II circumaural headphones. For the vowel and consonant tests, a graphical user interface was used that enabled the subjects to indicate their response by clicking a button corresponding to the syllable played. For the sentence test, subjects were asked to write down the words they heard. The sentences were scored in terms of percent words identified correctly (all words were scored). During the practice session, the identity of the test syllables (vowels or consonants) and sentences was displayed on the screen.

At the beginning of each test the subjects were presented with a practice session in which the speech materials were processed through the same number of channels used in the test and presented monaurally in quiet and in +5 dB speech-shaped noise. For further practice, vowel and consonant tests were administered with feedback to each subject. Three repetitions were used in the feedback session. The practice session lasted approximately 2 h. After the practice and feedback sessions, the subjects were tested with the various dichotic and monaural conditions. The vowels and consonants were completely randomized and presented to the listeners six times. No feedback was provided during the test, and all the tests were done with speech embedded in +5 dB speech-shaped noise taken from the HINT database.

Two different dichotic conditions were considered. In the first condition, which we refer to as low–high dichotic condition, the low-frequency information (consisting of half of the total number of channels) was presented to one ear, and the high-frequency information (consisting of the remaining half high-frequency channels) was presented to the other ear. In the second dichotic condition, which we call odd–even (or interleaved) dichotic condition, the odd-index frequency channels were presented to one ear, while the even-index channels were presented to the other ear. In the monaural condition, the signal was presented monaurally to either the left or the right ear (chosen randomly) of the subject. The order in which the conditions and number of channels was presented was partially counterbalanced between subjects to avoid order effects. In the vowel and consonant tests, there were six repetitions of each vowel and each consonant. The vowels and the consonants were completely randomized. A different set of 20-sentence lists was used for each condition.

Pilot data showed that the 12-channel condition in +5 dB S/N yielded performance close to ceiling. Hence, for the 12-channel condition, we performed additional listening experiments to assess whether subjects were indeed integrating the information from the two ears, or whether they were receiving sufficient information in each of the two ears alone. For comparison with the odd–even stimuli presented dichotically, two additional conditions were created. In the first condition, the odd-index channels were presented to the left ear alone, and in the second condition, the even-index channels were presented to the right ear alone. Similarly, for comparison with the low–high stimuli presented dichotically, the low-frequency channels (lower half number of channels) were presented to the left ear alone and the high-frequency channels were presented to the right ear alone. Sentences, vowels and consonants were processed though the one-ear conditions for the odd-even condition comparison. Vowels and consonants were also processed through the one-ear conditions for the low–high comparison. No sentences were processed for the low–high comparison since there were an insufficient number of unique sentences in the HINT database.

B. Results

The mean percent correct scores for sentence, vowel and consonant recognition is shown in Fig. 1 as a function of number of channels for different presentation modes.

1. Sentences

A two-way ANOVA with repeated measures, using spectral resolution (number of channels) and presentation mode (monaural, dichotic odd–even and dichotic low–high) as within-subject factors, showed a significant main effect of spectral resolution [$F(2,16)=82.05$, $p<0.0005$], a significant effect of presentation mode [$F(2,16)=10.57$, $p=0.001$], and a significant interaction [$F(4,32)=3.36$, $p=0.02$] between spectral resolution and presentation mode. *Posthoc* tests according to Tukey (at alpha$=0.05$) showed that there was a significant ($p<0.05$) difference between the performance obtained with the two dichotic conditions for 12 and 8 channels, but not for 6 channels. There was also a significant difference ($p<0.05$) between the performance obtained with dichotic presentation (low-high) and monaural presentation for the 6- and 8-channel conditions, but not for the 12-channel condition.

Figure 2 (top panel) compares the dichotic performance (12 channels) obtained on sentence recognition with the performance obtained when the even channels were presented to the left ear alone, and the odd channels were presented to the right ear alone. *Posthoc* Fisher's LSD tests showed that there was a significant difference ($p<0.05$) between the one-ear performance and the dichotic performance on sentence recognition, suggesting that subjects were able to integrate the information from the two ears. That is, the information presented in each ear alone was not sufficient to recognize sentences in +5 dB S/N with high (>90%) accuracy.

2. Vowels

A two-way ANOVA with repeated measures showed a significant main effect of spectral resolution [$F(2,16)=46.05$, $p<0.0005$], a significant effect of presentation mode [$F(2,16)=4.79$, $p=0.023$], and a significant interaction [$F(4,32)=7.16$, $p<0.0005$] between spectral resolution and presentation mode on vowel recognition. *Posthoc* tests

according to Tukey showed that there was a significant ($p < 0.05$) difference between each of the dichotic conditions and the monaural condition for eight-channels. There was also a significant difference between the two dichotic conditions for six channels.

Figure 2 compares the performance obtained dichotically (for both conditions) with the performance obtained with the left and right ears only. Posthoc Fisher's LSD tests showed that there was a significant difference ($p < 0.05$) between the one-ear performance and the dichotic performance on vowel recognition, suggesting that subjects were able to integrate the information from the two ears and obtain a vowel score higher than the score obtained with either ear alone.

3. Consonants

A two-way ANOVA with repeated measures showed a significant main effect of spectral resolution [$F(2,16) = 46.05$, $p < 0.0005$], a nonsignificant effect of presentation mode [$F(2,16) = 4.79$, $p = 0.34$], and a significant interaction [$F(4,32) = 7.16$, $p = 0.002$] between spectral resolution and presentation mode on consonant recognition. As shown in Fig. 1, the presentation mode (dichotic versus monaural) had no effect on consonant recognition.

Figure 2 compares the consonant performance obtained dichotically against the performance obtained with either the left or right ears alone. Posthoc Fisher's LSD tests showed that there was a significant difference ($p < 0.05$) between the one-ear performance and the dichotic performance on consonant recognition.

The consonant confusion matrices were also analyzed in terms of percent information transmitted as per Miller and Nicely (1955). The feature analysis is shown in Fig. 3. A two-way ANOVA performed on each feature separately showed a marginally significant effect of presentation mode for manner ($p = 0.04$) and voicing ($p = 0.047$) and a nonsignificant effect for place ($p = 0.31$). There was a significant interaction for place and voicing ($p < 0.005$), but not for manner ($p = 0.13$). There was significant effect ($p < 0.005$) of spectral resolution for all three features.

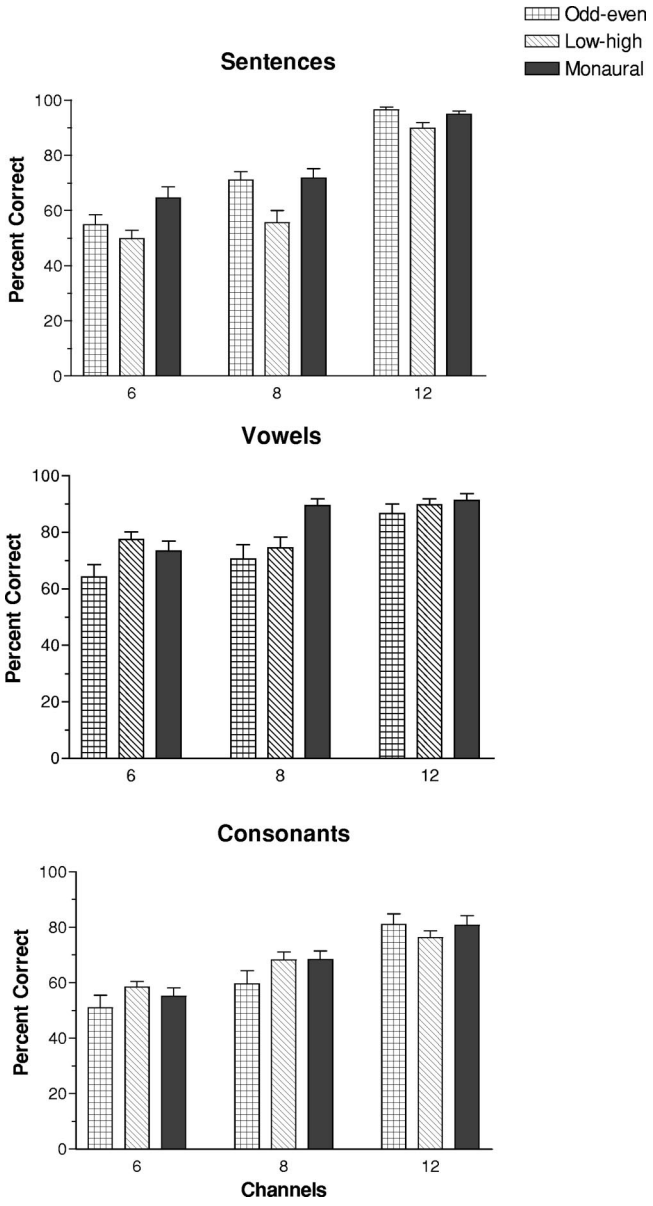


FIG. 1. Mean identification of sentences (scored in percent words correct), vowels, and consonants as a function of number of channels and presentation mode (monaural versus dichotic). Error bars indicate standard errors of the mean.

C. Discussion

The above results indicate that the effect of presentation mode (dichotic versus monaural) differed across speech materials and degree of spectral resolution. The recognition of sentences was affected the most. Recognition of vowels was also affected but to a lesser degree. As it is evident from the mean performance on sentences processed through six and eight channels, subjects were not able to fuse sentences presented dichotically in the low–high condition with the same accuracy as when presented monaurally. Subjects were also not able to fuse vowels, processed through eight channels and presented dichotically (in either low/high or even–odd conditions), with the same accuracy as when presented monaurally. It should be noted that there was large subject variability in performance (see Fig. 4) for vowels and sentences processed through six channels (see, for instance, subjects' S1, S2 and S7 performance on sentence recognition, and subjects' S4, S5 and S7 performance on vowel recognition). In contrast, subjects were able to fuse vowels and sentences processed through 12 channels very accurately and more consistently. These results suggest that spectral resolution may have a significant effect on spectral fusion depending on the dichotic presentation mode (low/high versus interleaved). No such effect was found for consonants. Consonants were fused accurately with both dichotic presentations regardless of the spectral resolution. The difference in effect on spectral fusion between consonants and the other speech materials suggests that vowels and sentences might be better (more sensitive) speech materials to use in studies of dichotic speech perception. This conclusion must be viewed with caution, taking into account the fact the performance variability might be partially due to the variability in speech material used in this study, with the vowel material being more variable (produced by both male and female talkers, with mul-

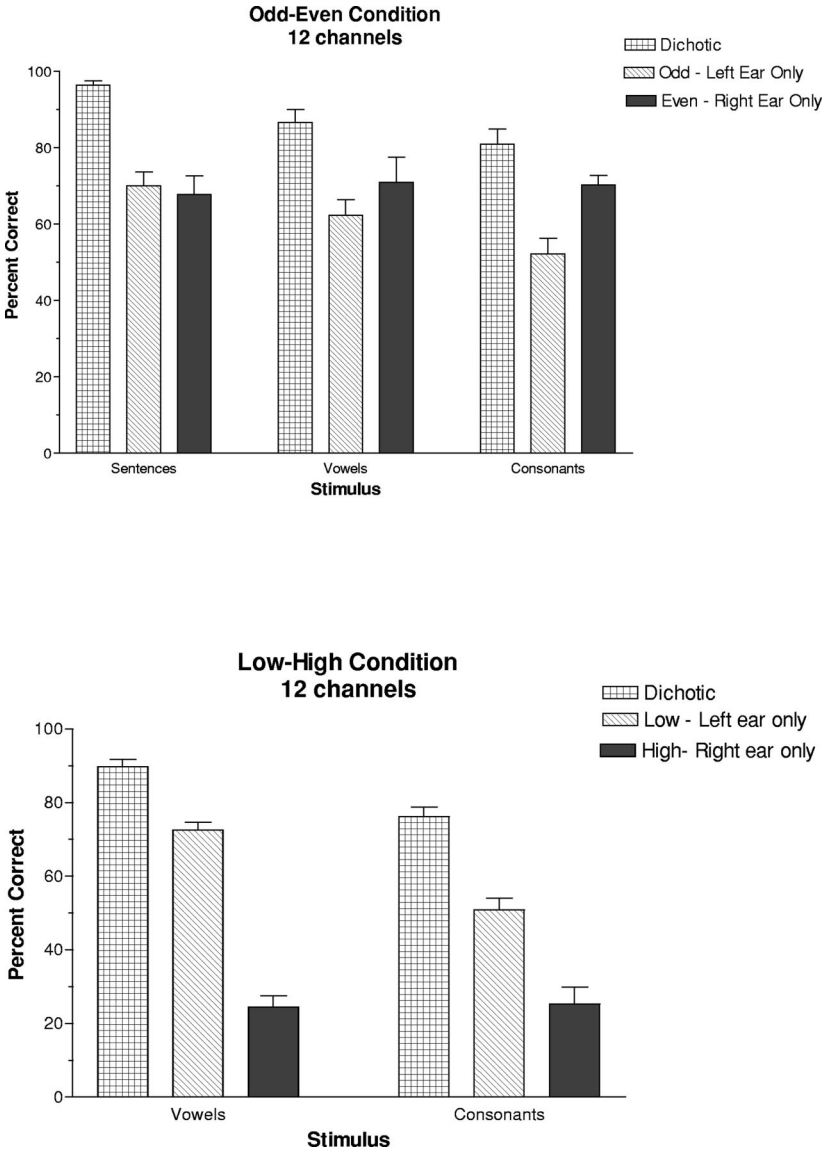


FIG. 2. (Top panel) Mean speech recognition obtained when the odd- and even-index channels were presented dichotically (hatched bars), the odd-channels were presented to the left ear only (diagonally filled bars), and the even-channels were presented to the right ear only (dark bars). (Bottom panel). Mean vowel and consonant recognition obtained when the low- and high-frequency channels were presented dichotically (hatched bars), the low-frequency channels were presented to the left ear only (diagonally filled bars), and the high-frequency channels were presented to the right ear only (dark bars). Speech materials were processed through 12 channels. Error bars indicate standard errors of the mean.

tiple productions of each vowel) and the consonant material being less variable (produced by a single male talker with a single production of each consonant).

When interpreting the results on vowel recognition, two confounding factors need to be considered. First, the filter spacing was logarithmic for the 6-channel condition and mel-like for the 8- and 12-channel conditions. We cannot exclude the possibility that a different outcome might have been obtained had a different filter spacing was used, and this warrants further investigation. Second, the 8- and 12-channel conditions had a slightly larger (by about 38–108 Hz) overall bandwidth than the 6-channel condition. Given that the roll-off of the sixth-order filters used was not too sharp, and the results from our previous study (Loizou *et al.*, 2000a) indicate no significant difference on vowel recognition between different signal bandwidths, we do not believe that the additional bandwidth in the 8- and 12-channel conditions affected the outcome of this study.

No effect on spectral fusion was found in this study in the identification of consonants processed through a small number of channels. This outcome is consistent with the findings in the study reported by Lawson *et al.* (1999) with bilateral cochlear implant listeners. Lawson *et al.* presented consonants in /aCa/ context to two bilateral implant users. The consonants were presented dichotically in the same two conditions used in our study, even–odd and low–high conditions. The bilateral subjects were fitted with six- and eight-channel processors. Results showed a small advantage of dichotic stimulation, however the difference was not statistically significant. No experiments were done in Lawson *et al.* (1999) with other speech materials. Identification of consonants, in general, is known to be robust to extremely low spectral-resolution conditions (Shannon *et al.*, 1995; Dorman *et al.*, 1997) and even conditions in which large segments of the consonant spectra are missing (Lippmann, 1996; Kasturi *et al.*, 2002). Dichotic identification of consonants does not seem to be an exception.

The fact that the listeners were not able to fuse with high accuracy sentences processed through six and eight channels and presented in the low–high dichotic condition cannot be easily explained given that we did not manipulate in this study the relative onset time, the F0, or the relative intensity of the two signals presented to each ear. One possible explanation is that the information presented in each ear was so

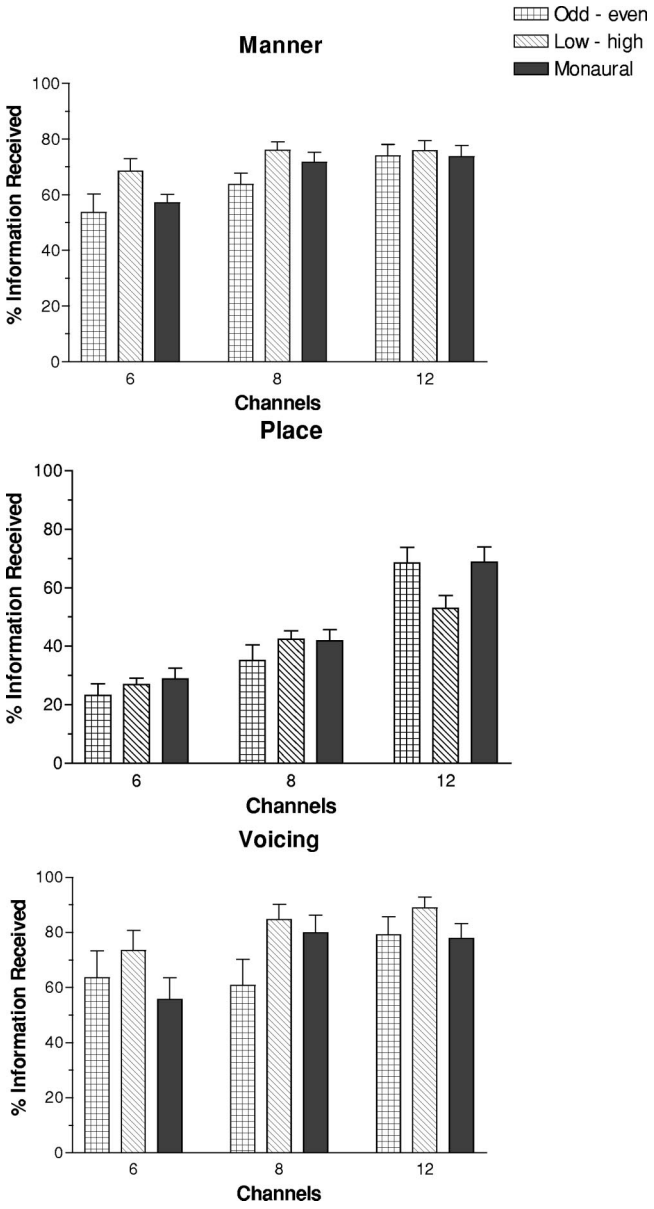


FIG. 3. Mean identification of the consonant features ''manner,'' ''place,'' and ''voicing'' as a function of number of channels and presentation mode (monaural versus dichotic). Error bars indicate standard errors of the mean.

impoverished, due to the high noise level and low spectral resolution (three or four channels in each ear), that it was not perceived as being speech by the auditory system, hence it was not integrated centrally. This is based on the assumption that the ability to fuse a sound presented to one ear with another sound presented to the other ear must depend on recognizing that the two components come from the same speech utterance (and probably the same talker). In contrast, in the 12-channel condition the subjects had access to 6 channels of information in each ear, and we know from Fig. 2 that moderate levels of performance can be achieved in +5 dB S/N with only 6 channels presented to each ear. A similar outcome was also reported by Dorman *et al.* (2001) with HINT sentences presented dichotically in quiet with even channels fed in one ear and odd channels fed in the other ear. Sentences were processed through six channels in a manner similar to this study. Sentence scores obtained monaurally were significantly higher than the scores obtained dichotically.

In addition to observing a significant difference between the dichotic and the monaural conditions, a significant difference was also observed between the two dichotic conditions. The way spectral information was split and presented to the two ears was important for vowel and sentence recognition, but not for consonant recognition. For sentences processed through 8 and 12 channels, the mean scores obtained with the interleaved (odd–even) condition were significantly higher than the scores obtained with the low–high condition. For vowels processed through six channels, the scores obtained with the low–high condition were significantly higher than the scores obtained with the interleaved condition. The higher scores obtained with the low–high condition in vowel recognition may be attributed to the fact that in this condition F1 information (low-frequency information) was presented to one ear and F2 information (higher-frequency information) was presented to the other ear. This condition must be easier to deal with compared to the more challenging condition (interleaved) in which pieces of F1 and F2 information are distributed between the two ears. Subjects did not have difficulty piecing together the F1/F2 information, however, when the vowels were processed through 8 and 12 channels. The above explanation can not be easily extended to sentences, because the situation with sentences differs from that of vowels, in that listeners are relying on other cues, besides F1/F2 information, for word recognition.

One possible explanation for the higher scores obtained in sentence recognition with the interleaved condition is that it provides greater dichotic release from masking [a phenomenon first reported by Rand (1974)] compared to the low–high condition. Rand (1974) presented the F1 of CV syllables to one ear, and the F2 attenuated by 40 dB to the other ear, and observed that subjects were able to identify the consonants accurately despite the large relative intensity differences between the two formants. However, when he attenuated the upper formants by 30 dB and presented the stimuli to both ears (i.e., diotically), the listeners were unable to identify the consonants accurately. He attributed the advantage of dichotic presentation to release from spectral masking. Others (e.g., Lunner *et al.*, 1993; Pandey *et al.*, 2001; Lyregaard, 1982; Franklin, 1975) have attempted to exploit the dichotic release from masking in bilateral hearing aids and advocated the use of dichotic presentation as a means of compensating for the poor frequency selectivity of listeners with sensorineural hearing loss. Lunner *et al.* (1993) fitted three hearing-impaired listeners with eight-channel hearing aids and presented sentences dichotically, with the odd-number channels fed to one ear and the even-number channels fed to the other ear. An improvement of 2 dB in speech reception threshold (SRT) was found compared to the condition in which the sentences were presented binaurally. Franklin (1975) investigated the effect of presenting a low-frequency band (240–480 Hz) and a high-frequency band (1020–2040 Hz) on consonant recognition in six hearing-impaired listeners with moderate to severe sensorineural hearing loss. The scores obtained when the low- and high-frequency bands were presented to opposite ears were sig-

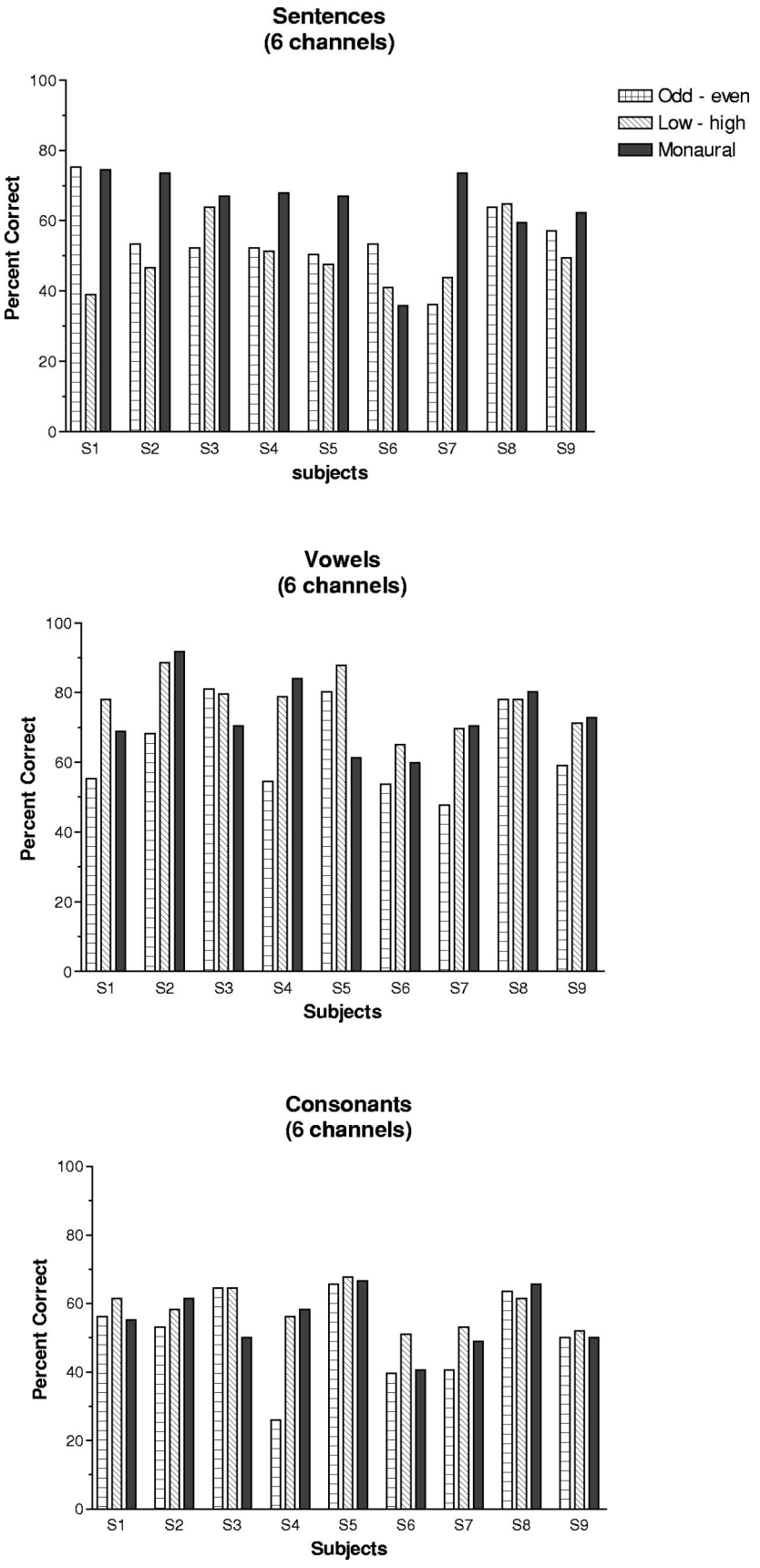


FIG. 4. Individual subject performance for speech materials processed through six channels.

nificantly higher than the scores obtained when the two bands were presented to the same ear. Unlike the above two studies, a few other studies (e.g., Lyregaard, 1982; Turek *et al.*, 1980) found no benefit of dichotic presentation with hearing-impaired listeners tested on synthetic /b d g/ identification (Turek *et al.*, 1980) or vowel/consonant identification (Lyregaard, 1982).

In the present study, speech was synthesized as a sum of

a small number (6–12) of sinewaves. Sinewave speech was advocated in our earlier work (Dorman *et al.*, 1997) as an alternative [to noise-band simulations (Shannon *et al.*, 1995)] acoustic model for cochlear implants. This acoustic model is not meant by any means to mimic the percept elicited by electrical stimulation, but rather to serve as a tool in assessing CI listener's performance in the absence of confounding factors (e.g., neuron survival, electrode insertion depth, etc.) associated with cochlear implants. We previously demonstrated on tests of reduced intensity resolution (Loizou *et al.*, 2000b), reduced spectral contrast (Loizou and Poroy, 2001), and number of channels of stimulation (Dorman and Loizou, 1998) that the performance of normal-hearing subjects, when listening to speech processed through the sinewave acoustic model, is similar to that of, at least, the better performing patients with cochlear implants. In that context, the results of the present study might provide valuable insights to bilateral cochlear implants. The performance obtained with the interleaved dichotic condition was not significantly different from the performance obtained monaurally when the speech materials were processed through 12 channels. For bilateral patients receiving a large number of channels (>12), dichotic stimulation might therefore provide an advantage over unilateral stimulation, in that it can potentially reduce possible channel interactions since the electrodes can be stimulated in an interleaved fashion across the ears without sacrificing performance. For patients receiving only a small number of channels of information, dichotic (electric) stimulation might not produce the same level of performance as unilateral (monaural) stimulation, at least for sentences presented in the low–high dichotic condition. The large variability in performance among subjects should be noted however (Fig. 4). Lastly, of the two methods that can be used to present spectral information dichotically, the interleaved method is recommended since it consistently outperformed the low–high method in sentence recognition.

III. SUMMARY AND CONCLUSIONS

This study investigated the hypothesis that spectral resolution might be a factor, in addition to relative onset time, influencing spectral fusion when the spectral information is split and presented dichotically to the two ears. Two different methods of splitting the spectral information were investigated. In the first method, which we refer to as the odd–even method, the odd-index channels were presented to one ear and the even-index channels to the other ear. In the second method, which we refer to as low–high method, the lower-frequency channels were presented to one ear and the high-frequency channels to the other ear. The results of the present study suggest the following.

(i) Spectral resolution did affect spectral fusion, and the effect differed across speech materials and the two dichotic presentation modes considered. The recognition of sentences presented in the low–high dichotic condition was affected the most. Sentences, processed through six and eight channels and presented dichotically in the low–high condition, were not fused as accurately as when presented monaurally. Vowels processed through eight channels and presented dichotically (in either the interleaved or low–high conditions) were not fused as accurately as when presented monaurally. In contrast, vowels and sentences processed through 12 channels were fused very accurately.

(ii) Dichotically presented consonants were fused as well as monaurally presented consonants independent of spectral resolution (number of channels) and the way spectral information was split and presented to the two ears. That is, the performance obtained monaurally was the same as the performance obtained dichotically in all conditions. The difference in outcomes between the consonants and the other speech materials suggests that perhaps the vowels and sentence materials are better (more sensitive) materials to be used in studies of dichotic speech recognition assessment.

(iii) A significant difference in performance was found in vowel and sentence recognition between the two methods used to split the spectral information for dichotic presentation. For sentence recognition, the scores obtained with the odd–even (interleaved) dichotic method were significantly higher than the scores obtained with the low–high dichotic method. One possible explanation for this difference is that the odd–even method provides a greater dichotic release from masking compared to the low-high method.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Richard Tyler and Dr. Ken Grant for providing valuable comments to this paper. This research was supported by Grant No. R01 DC03421 from the National Institute of Deafness and other Communication Disorders, NIH. This project was the basis for the Master's thesis of the second author (AM) in the Department of Electrical Engineering at the University of Texas—Dallas.

Broadbent, D. E., and Ladefoged, P. (**1957**). ''On the fusion of sounds reaching different sense organs,'' J. Acoust. Soc. Am. **29**, 708–710.

Cherry, E. C. (**1953**). ''Some experiments on the recognition of speech, with one and with two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Cutting, J. E. (**1976**). ''Auditory and linguistic process in speech perception: Inferences from six fusions in dichotic listening,'' Psychol. Rev. **83**, 114–140.

Darwin, C. J. (**1981**). ''Perceptual grouping of speech components differing in fundamental frequency and onset-time,'' Q. J. Exp. Psychol. A **33A**, 185–207.

Dorman, M., and Loizou, P. (**1998**). ''The identification of consonants and vowels by cochlear implants patients using a 6-channel CIS processor and by normal hearing listeners using simulations of processors with two to nine channels,'' Ear Hear. **19**, 162–166.

Dorman, M., Loizou, P., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Dorman, M., Loizou, P., Spahr, A. J., Maloff, E. S., and Wie, S. V. (**2001**). ''Speech understanding with dichotic presentation of channels: Results from acoustic models of bilateral cochlear implants,'' 2001 Conference on Implantable Auditory Prostheses, Asilomar, Monterey, CA.

Franklin, B. (**1975**). ''The effect of combining low and high-frequency bands on consonant recognition in the hearing impaired,'' J. Speech Hear. Res. **18**, 719–727.

Halwes, T. G. (**1970**). ''Effects of dichotic fusion on the perception of Speech,'' Doctoral dissertation, University of Minnesota, Dissertation Abstracts International, 31, p. 1565B, (University Microfilms No. 70-15, 736).
Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Kasturi, K., Loizou, P., Dorman, M. (**2002**). ''The intelligibility of speech with holes in the spectrum,'' J. Acoust. Soc. Am. **112**, 1102–1111.
Lawson, D., Wilson, B., Zerbi, M., and Finley C. (**1999**). ''Speech processors for auditory prostheses,'' *Fourth Quarterly Progress Report*, Center for Auditory Prosthesis Research Triangle Institute.
Lippmann, R. P. (**1996**). ''Accurate consonant perception without mid-frequency speech energy,'' IEEE Trans. Speech Audio Process. **4**(1), 66–69.
Loizou, P., and Poroy, O. (**2001**). ''Minimum spectral contrast needed for vowel identification by normal-hearing and cochlear implant listeners,'' J. Acoust. Soc. Am. **110**, 1619–1627.
Loizou, P., Dorman, M., and Powell, V. (**1998**). ''The recognition of vowels produced by men, women, boys and girls by cochlear implant patients using a six-channel CIS processor,'' J. Acoust. Soc. Am. **103**, 1141–1149.
Loizou, P., Dorman, M., and Tu, Z. (**1999**). ''On the number of channels needed to understand speech,'' J. Acoust. Soc. Am. **106**, 2097–2103.
Loizou, P., Poroy, O., and Dorman, M. (**2000a**). ''The effect of parametric variations of cochlear implant processors on speech understanding,'' J. Acoust. Soc. Am. **108**, 790–802.
Loizou, P., Dorman, M., Poroy, O., and Spahr, T. (**2000b**). ''Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution,'' J. Acoust. Soc. Am. **108**, 2377–2388.
Lunner, T., Arlinger, S., and Hellgren, J. (**1993**). ''8-channel digital filter bank for hearing aid use: Preliminary results in monaural, diotic and dichotic modes,'' Scand. Audiol. Suppl. **38**, 75–81.
Lyregaard, P. E. (**1982**). ''Frequency selectivity and speech intelligibility in noise,'' Scand. Audiol. Suppl. **15**, 113–122.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Nilsson, M., Soli, S., and Sullivan, J. (**1994**). ''Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,'' J. Acoust. Soc. Am. **95**, 1085–1099.
Pandey, P. C., Jangamashetti, D. S., and Cheeran, A. N., (**2001**). ''Binaural dichotic presentation to reduce the effect of temporal and spectral masking in sensorineural hearing impairment,'' 142nd Meeting of the Acoustical Society of America, Ft. Lauderdale, FL.
Rand, T. C. (**1974**). ''Dichotic release from masking for speech,'' J. Acoust. Soc. Am. **55**, 678–680.
Shannon, R., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Turek, S., Dorman, M., Franks, J., and Summerfield, Q. (**1980**). ''Identification of synthetic /bdg/ by hearing impaired listeners under monotic and dichotic formant presentation,'' J. Acoust. Soc. Am. **67**, 1031–1040.
Tyler, R., Preece, J., and Lowder, M. (**1987**). ''The Iowa audiovisual speech perception laser videodisc,'' Laser Videodisc and Laboratory Report, Dept. of Otolaryngology, Head and Neck Surgery, University of Iowa Hospital and Clinics, Iowa City.

The effect of speech presentation level on measurement of auditory acclimatization to amplified speech[a)]

Kevin J. Munro[b)] and Mark E. Lutman
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 3 May 2002; revised 30 March 2003; accepted 7 April 2003)

A systematic improvement in auditory performance over time, following a change in the acoustic information available to the listener (that cannot be attributed to task, procedural or training effects) is known as auditory acclimatization. However, there is conflicting evidence concerning the existence of auditory acclimatization; some studies show an improvement in performance over time while other studies show no change. In an attempt to resolve this conflict, speech recognition abilities of 16 subjects with bilateral sensorineural hearing impairments were measured over a 12-week period following provision of a monaural hearing instrument for the first time. The not-fitted ear was used as the control. Three presentation levels were used representing quiet, normal, and raised speech. The results confirm the presence of acclimatization. In addition, the results show that acclimatization is evident at the higher presentation levels but not at the lowest. [DOI: 10.1121/1.1577556]

PACS numbers: 43.71.Ky, 43.71.Pc, 43.66.Ts [CWT]

I. INTRODUCTION

With the provision of amplification, the hearing-impaired listener receives newly available speech cues that were previously inaudible. In addition, cues that were previously audible are shifted towards the higher intensity end of the auditory neural representation. These changes may immediately confer greater intelligibility. However, provision of amplification may result in additional improvement in performance over time, possibly as a result of reorganization within the auditory cortex. This improvement over time is known as auditory acclimatization and has been defined by Arlinger *et al.* (1996) as ''A systematic improvement in auditory performance over time, following a change in the acoustic information available to the listener. It involves an improvement in performance that cannot be attributed to task, procedural or training effects.'' Evidence for auditory acclimatization after prolonged use of acoustic hearing aids is mixed; some authors state the acclimatization effects are small or nonexistent (Turner and Bentler, 1998). For a review, see the articles by Turner *et al.* (1996) and Palmer *et al.* (1998).

The most frequently cited acoustic hearing instrument studies providing evidence of auditory acclimatization include Cox and Alexander (1992), Cox *et al.* (1996), Gatehouse (1992, 1993), and Horwitz and Turner (1997). Cox and Alexander (1992) studied eight new hearing instrument users and four experienced users fitted with new instruments. Speech recognition was measured using the Connected Speech Test (CST) at the time of fitting and 10 weeks later. No significant change in mean benefit (aided minus unaided performance) was measured in noisy or reverberant listening conditions. However, in a low-noise background there was a statistically significant increase in mean benefit over time of 5%–6%. There was no control group, so it is not clear if the improvement was due to practice from repeated use of the same test material. In a further study, Cox *et al.* (1996) included experienced hearing instrument users as a control. The experimental group of 22 new hearing instrument users was fitted monaurally and tested at the time of fitting and 12 weeks later. The mean benefit on CST increased from 4% at the time of fitting to 8% 12 weeks later; there was no improvement in the control group. The change in benefit was due to an increase in aided performance with no change in unaided performance.

In 1992, Gatehouse used the Four-Alternative Auditory Feature (FAAF) test to track performance over a 12-week postfitting period in four new hearing instrument users fitted monaurally. Benefit increased from 5% at the time of fitting to greater than 15% at 12 weeks postfitting; the improvement commenced at around 6 weeks postfitting and was due to both an increase in the aided condition and a decrease in the unaided condition. No improvement in benefit was observed in the not-fitted (control) ear of these monaurally aided subjects. In a subsequent experiment, Gatehouse (1993) refitted 36 experienced hearing instrument users with a new instrument that provided greater high-frequency amplification than their previous one. Aided performance was measured on the FAAF test at 0, 8, and 16 weeks after the new fitting. Mean scores with the old and new hearing instrument were similar initially but increased significantly by 2.3% at 8 weeks and 4.4% at 16 weeks.

Horwitz and Turner (1997) compared 13 new hearing instrument users fitted monaurally with 13 experienced hearing instrument users as the controls. Speech recognition was measured over 18 weeks using the Nonsense Syllable Test (NST) with the hearing instrument at a fixed initial gain set-

[a)]Portions of this work were presented at the British Society of Audiology Experimental Short Papers Meeting, Oxford, UK, September 2001.

[b)]Now at: Human Communication and Deafness Group, School of Education, University of Manchester, Manchester M13 9PL, UK. Electronic mail: kevin.j.munro@man.ac.uk

ting and at the subject's daily adjusted gain setting. There was a gradual increase in mean benefit from around 6% at the time of fitting to around 14% at 18 weeks for the fixed gain setting in the new users. This was due to increases in aided performance rather than reductions in the unaided performance. A similar finding was reported for the user-adjusted gain setting. There was no improvement in the control group.

By contrast, several studies have failed to demonstrate auditory acclimatization, including Bentler *et al.* (1993), Humes *et al.* (1996, 2002), Munro and Lutman (2000), Saunders and Cienkowski (1997), Surr *et al.* (1998), and Taylor (1993). There are at least three possible explanations for this. First, the subjects may have had little opportunity to improve over time due to mild hearing loss (Taylor, 1993), previous experience of a hearing instrument (Bentler *et al.*, 1993), or limited use (Taylor, 1993; Bentler *et al.*, 1993). Second, Robinson and Summerfield (1996) suggest that the negative findings may be due to the specificity of perceptual learning: the test methods employed may not have picked up changes that had occurred. Third, findings may relate to the presentation level of the test material used to demonstrate acclimatization. The main focus of the present study is this third possibility.

Munro and Lutman (2000) used very similar methodology to Gatehouse (1992), but did not show an improvement in benefit over time. One difference between the two studies is that Munro and Lutman used a presentation level of 58 dB SPL, whereas Gatehouse used a presentation level of 65 dB SPL. This difference in presentation level is reflected in a difference in the mean initial benefit score between the studies (14% for Munro and Lutman but only 5% for Gatehouse); at higher presentation levels the unaided speech is already audible so the difference between the aided and unaided score is small. The dependence of acclimatization on presentation level can be rationalized as follows. Consider a subject who experiences speech in everyday life over the range 55–75 dB SPL before fitting and 75–95 dB SPL after fitting (i.e., the hearing instrument provides 20 dB gain). For simplicity, the shape of the gain/frequency response curve is not considered here. If speech test materials are presented at 55 dB, they are amplified to 75 dB and hence there is little difference compared to that experienced before amplification. However, if material is presented at 70 dB and amplified to 90 dB, this reaches a level not previously experienced for normal speech. The organism must adapt and reorganize to discriminate and utilize the speech cues that are now coded into an unfamiliar part of the neural representation of sound. It is proposed that this process of reorganization underlies the phenomenon of auditory acclimatization. In the foregoing example, reorganization would be required to discriminate speech presented at 70 dB but not at 55 dB.

The aim of this study is to test the hypothesis that auditory acclimatization is only revealed when testing involves using speech materials at higher presentation levels. This is achieved by measuring aided and unaided speech recognition performance in newly aided subjects at three different speech presentation levels repeatedly over 12-weeks postfitting. The null hypothesis is that there are no differential changes in benefit over time as a function of presentation level. The alternative hypothesis is that improvements are greater at the higher presentation levels and negligible at the lowest level. In order to remove hearing instrument gain setting as a confounding variable, subjects are fitted with linear amplification with the gain control disabled for the duration of the study.

II. METHODS

A. Subjects

Sixteen subjects were recruited (10 male, 6 female) with a mean age of 70 years (s.d.$\pm$5.5). The sample size was calculated for paired data with a mean difference of 4% and a standard deviation of the difference of 5%. Fifteen subjects were required for a statistical power of 80% at a two-tailed significance level of 5% using Student's *t*-test, but 16 were recruited to allow for attrition. No subjects withdrew from the study and no subjects missed any test session.

Subjects were all first-time hearing instrument users recruited from the local hospital audiology service. They complained of hearing disability, and pure-tone audiometry revealed a symmetrical, mild-to-moderate, sloping, high-frequency sensorineural hearing impairment. Exclusion criteria included an asymmetry in air conduction thresholds of greater than 15 dB at two or more frequencies, an air–bone gap greater than 15 dB at any test frequency, and abnormal middle-ear function assessed using oto-admittance audiometry. Prior to hearing instrument management, subjects were interviewed and informed that the aim of the experiment was to investigate the benefit provided by a single hearing instrument; however, they were naive to the changes expected over time. The study was approved by local hospital and university ethical committees.

Table I summarizes hearing thresholds and uncomfortable loudness levels, showing symmetrical, high-frequency sensorineural hearing impairments with reduced dynamic range. The mean change in hearing threshold level between the beginning and end of the study [initial minus final measurement] was 0 dB (s.d.$\pm$6 dB). Over 80% of hearing thresholds changed by less than 5 dB at each audiometric frequency and 90% changed by less than 10 dB; this is consistent with audiometric test–retest differences reported in the literature (for example, Robinson, 1991).

Each subject received standard shell earmolds to which a parallel 0.8-mm vent was added. This is standard clinical practice to equalize pressure and avoid a build-up of moisture in the ear canal: it is likely to have a negligible effect on the low-frequency gain of the amplified sound path. The same model of hearing instrument was used for all subjects: Phonak Sono-Forte 331X-L PiCS which is a miniature, high-gain, digitally programmable BTE hearing instrument. It has a three-way audio filter allowing considerable flexibility when tailoring the frequency response. It can also store three programs that can only be accessed using a hand control. The subjects were not issued the hand control and were therefore unable to change programs during the study. The fittings were all monaural, which allowed the not-fitted ear to be used as a control. The not-fitted control ear underwent the

TABLE I. Summary of audiometric data for the subjects. The table includes the mean air conduction, not-masked bone conduction hearing threshold levels, and mean uncomfortable loudness levels. One standard deviation is given in brackets ($n=16$).

Frequency (kHz)	0.25	0.5	0.75	1	1.5	2	3	4	6
				Air conduction (dB HL)					
Fitted ear	33 (12)	34 (16)	38 (14)	44 (12)	50 (10)	53 (8)	60 (9)	64 (6)	71 (9)
Control ear	34 (11)	36 (13)	38 (13)	44 (11)	49 (9)	52 (11)	58 (10)	62 (8)	67 (12)
				Bone conduction (dB HL)					
Not-masked		30 (13)		35 (8)		50 (11)		55 (7)	
				Uncomfortable loudness level (dB HL)					
Fitted ear		102 (11)		101 (10)		104 (12)		108 (14)	
Control ear		101 (10)		100 (11)		105 (11)		111 (13)	

same test protocol as the fitted ear using the subjects' own hearing instrument. The only difference between test and control ears was wearing of the instrument between sessions. It was important to have a control condition because increases in performance might otherwise be explained by practice effects due to repeated exposure to the test material. However, since both ears had identical opportunities for practice with the test materials, taking the difference between scores on the two ears gave a robust measure of acclimatization. Some subjects expressed a clear preference for ear to be fitted: the fitted ear in the remaining subjects was selected at random. Six subjects were fitted in the right ear and ten subjects were fitted in the left ear. The earmold for the ear to be the control ear was retained in the laboratory and not given to the subject.

The hearing instrument response was tailored to the NAL-RM target values (Byrne and Dillon, 1986) using the real-ear insertion gain (REIG) protocol on the Rastronics Portarem 2000 probe-tube measurement system. This involved measuring the real-ear unaided response and subtracting it from the real-ear aided response for each subject. Table II shows the mean target values (± 1 s.d.) for the fitted ear along with the best match to target. The mean target gain from 1–4 kHz (where the hearing impairment is greatest) is 22–23 dB. There is very good agreement between the target values and the best match: the mean difference is typically less than 1–2 dB. Also shown is the user-gain setting that was fixed for the duration of the study. The subjects were free to adjust the gain control of the hearing instrument during the first few days after fitting before returning to the laboratory to have this disabled at their preferred user-gain setting. All adjustments were made within 7 days of fitting. The mean user gain was 4 dB below the target gain from 1–4 kHz. It is possible that subjects may have been conservative in selecting their preferred gain because they were aware that this would be fixed for the duration of the study. However, other studies have reported a similar reduction in user gain compared to NAL-RM target values (Humes *et al.*, 2000). Since the gain control was disabled for the duration of the study, special care was taken to avoid exceeding uncomfortable loudness levels. The maximum output was set close to uncomfortable loudness level and adjusted if the subject reported any undue discomfort. The mean real-ear saturation response was set around 104 dB SPL using mainly peak-clipping output limiting. These settings were entered into memory one of the hearing instruments since this was the default memory used by the subject in their home environment.

The REIG and the HA2 2-cc coupler gain were measured at the time of fitting and again at 6 and 12 weeks postfitting. The results are shown in Table III. The mean change in gain was less than 1 dB. The standard deviation was higher for insertion gain than for 2-cc coupler gain (typically 2.5 and 1.5 dB, respectively). While it is possible that there were real changes in REIG, the most likely explanation is that probe-tube measurements are inherently more variable than coupler measurements. There are no studies available that report long-term test–retest variability for REIG; however, the standard deviations in the present study are typical of those reported for short-term test–retest data by Hawkins *et al.* (1991). Thus, each subject was provided with relatively constant gain.

B. Test materials

The speech recognition test used was a digitized version of the Four-Alternative Auditory Feature (FAAF) test, as used in acclimatization studies of Gatehouse (1992, 1993). The FAAF test is a forced-choice word recognition task based on the rhyme test principle, described by Foster and Haggard (1979, 1987). It consists of 20 sets of four minimally paired words, each based on two binary auditory/

TABLE II. Mean real-ear insertion gain (dB) at NAL-RM target, best match to target and at the preferred user setting, as a function of frequency, in the test ear. One standard deviation is given in brackets.

Frequency (kHz)	0.25	0.5	0.75	1	1.5	2	3	4
Target (dB)	3 (3.4)	10 (6.6)	16 (5.7)	22 (5.1)	23 (4.2)	22 (3.2)	23 (2.5)	24 (2.7)
Match (dB)	3 (2.8)	8 (7.5)	16 (4.7)	21 (4.6)	24 (5.1)	24 (4.5)	23 (2.5)	22 (4.2)
User (dB)	1 (1.9)	3 (5.2)	12 (5.2)	17 (4.6)	20 (6.5)	20 (6.1)	19 (2.5)	17 (4.7)

TABLE III. Mean change in gain (dB) at the user setting, relative to the time of fitting, for the fitted ear. One standard deviation is given in brackets.

Frequency (kHz)	0.25	0.5	0.75	1	1.5	2	3	4
				Real-ear insertion gain				
Week 6 (dB)	0 (1.7)	0 (2.7)	0 (1.8)	0 (2.8)	0 (2.0)	0 (2.7)	0 (2.4)	1 (1.7)
Week 12 (dB)	0 (2.3)	0 (2.1)	1 (1.7)	0 (2.2)	1 (1.5)	0 (3.3)	−1 (2.7)	2 (3.6)
				2-cc coupler gain				
Week 6	0 (1.5)	0 (1.4)	−1 (1.2)	0 (1.2)	−1 (1.6)	−1 (1.4)	−1 (1.6)	−1 (1.5)
Week 12	1 (1.0)	2 (1.6)	0 (1.1)	0 (1.5)	0 (1.3)	−1 (1.6)	−1 (1.0)	−1 (0.9)

phonetic distinctions, giving an 80-item vocabulary. The items in a set differ on either initial or final consonant, and hence the test is particularly sensitive to high-frequency auditory capabilities. The key words occur in the context of the carrier phrase, ''Can you hear (*keyword*) clearly?'' One item is presented acoustically in the carrier phrase (spoken by a male speaker) and the subject's task is to select the correct word from the choice of four shown on a touch-screen monitor. The frequency spectrum of the key words is shown in Fig. 1 in comparison with the frequency response of conversational speech (from ANSI, 1997). The FAAF key words show slightly more emphasis at the high frequencies compared to the conversational speech.

The FAAF materials were replayed from a standard 16-bit computer sound card, at a sample rate of 20 kHz, routed via a Grason-Stadler GSI 61 clinical audiometer to a Fostex 6301B loudspeaker. They were presented against a background of steady noise that was filtered to give a similar long-term spectrum to key words and delivered by the same loudspeaker. The spectrum of this noise is also shown in Fig. 1. The loudspeaker was located at an azimuth of 0° and a distance of 1.5 m from the subject. Speech and noise levels were defined in terms of the overall SPL measured at the reference point, defined as the position occupied by the center of the subject's head, with the subject absent (ISO 8253-2, 1998).

Pearsons *et al.* (1977) measured the mean speech levels in a variety of settings and reported ''casual,'' ''normal,'' and ''raised'' speech for male speakers of 56, 61, and 68 dB SPL, respectively, with female speakers approximately 2 dB lower. Therefore, the FAAF test was presented at overall levels of 55, 62, and 69 dB SPL to approximate quiet, normal, and raised speech.

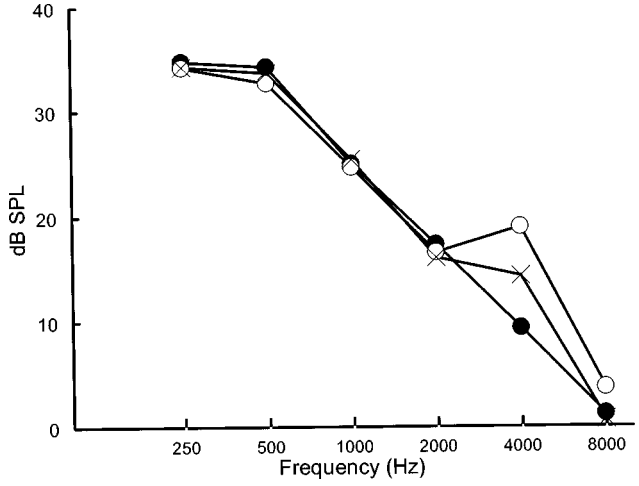


FIG. 1. Octave band levels for conversational speech at 62.35 dB SPL. FAAF words, open circles; FAAF-shaped noise, crosses; conversational speech from ANSI S3.7 (1997), filled circles.

The mean Speech Intelligibility Index (SII; ANSI, 1997) values for the aided and unaided condition were used to provide a guide to changes in audibility for each presentation level and are shown in Table IV. The equivalent speech spectrum level of the FAAF test was based on overall levels of 55, 62, and 69 dB SPL for the key words. The SII assumes that the speech area covers a dynamic range of 30 dB. The SII is the weighted sum of band audibility where the weight ranges from zero to 1 and represents the relative importance of each frequency band to the understanding of speech. For recognizing nonsense syllables, the band around 2 kHz is most important; for recognizing sentences, the band around 0.5 kHz is most important. The band importance function for the FAAF test has not been determined but is likely to be similar to that for nonsense syllables. The SII includes a level distortion factor (LDF) to account for the deterioration in speech performance that occurs when overall level exceeds 73 dB SPL. The equivalent noise spectrum of the FAAF-shaped noise spectrum was determined for each subject (based on the SNR). The aided values were obtained by

TABLE IV. Mean Speech Intelligibility Index for the fitted ear. The values in parentheses incorporate the desensitization correction from Pavlovic *et al.* (1986). The equivalent speech spectrum level of the FAAF test was based on overall SPLs of 55, 62, and 69 dB for the key words. The equivalent noise spectrum of the FAAF-shaped noise spectrum was determined for each subject (based on the SNR). The aided values were obtained by adding the user insertion gain to the speech and noise levels. The band importance function for nonsense syllables was used in the calculation.

	Overall
Unaided	
55 dB SPL	0.08(0.06)
62 dB SPL	0.12(0.09)
69 dB SPL	0.20(0.13)
Aided	
55 dB SPL	0.25(0.15)
62 dB SPL	0.39(0.21)
69 dB SPL	0.42(0.22)
Change (aided–unaided)	
55 dB SPL	0.17(0.09)
62 dB SPL	0.27(0.12)
69 dB SPL	0.22(0.09)

adding the user insertion gain to the speech and noise levels. [The values in parentheses in Table IV incorporate the desensitization correction from Pavlovic *et al.* (1986) to account for the poorer performance of hearing impaired subjects compared to normal-hearing subjects for a given level of audibility.] The low aided SII value at 55 dB SPL indicates that part of the speech envelope remains inaudible even after amplification. The small improvement after aiding at a presentation level of 69 dB SPL indicates that much of the speech signal was already audible before aiding. The SII values confirm that performance should improve as the presentation level increases, especially for the unaided condition.

Each subject was tested at a fixed signal-to-noise ratio (SNR) across all test conditions. A pilot study indicated that a single individually tailored SNR would give a score in the range 40%–80% across all test conditions which should be sufficiently high to maintain subject motivation but avoid ceiling effects. The SNR was determined for each subject using the adaptive FAAF test strategy described by Lutman and Clark (1986), which targets 71% correct with a speech presentation level of 62 dB SPL. SNR was defined as the difference in decibels between the SPL of the key words and the SPL of the noise. Subjects then practiced the test at all three presentation levels, while the investigator adjusted the SNR to compensate for any improvement in performance with practice; this normally involved increasing the noise by 1–3 dB over the course of several practice sessions. Performance was then measured throughout the study at this fixed SNR for all speech presentation levels. The SNR ranged from 0 to −5 dB across subjects with a median value of −2 dB. Subject performance was assessed as the number of words correctly identified and was expressed as percent correct. A random numbers table was employed to generate different orders of the 80 items for the multiple presentations involved.

Testing was carried out in a quasifree sound field [ISO 8253-2 (1998)] with reverberation times <0.3 s when measured in octave intervals from 0.125 to 4 kHz. The ambient noise levels measured at the reference test point (with the test equipment switched on but subject absent) was sufficiently low to have allowed measurement of binaural hearing thresholds in the sound field of 0 dB (maximum uncertainty +5 dB) from 0.25 to 4 kHz [ISO 8253-2 (1998)].

C. Procedures

Subjects attended several sessions (usually 4–6) each lasting approximately 2 h before obtaining their hearing instrument. This allowed tailoring of the hearing instrument frequency response and determination of the SNR required for presentation of the speech material. It also enabled the subject to become familiar with the testing procedure and materials. Once the subjects had completed these practice sessions, they were fitted with the hearing instrument for everyday use.

The baseline FAAF tests were performed within 7 days of fitting; testing could not commence until the fixed user gain had been established. Data for the initial test session is reported as occurring at zero days postfitting; however, for some subjects this may have taken place as late as 7 days postfitting. Testing was repeated at 6 and 12 weeks postfitting. The order of testing was balanced across subjects and test sessions.

Each test session commenced with a general discussion about use and progress with the hearing instrument since the previous visit. The subjects returned a diary that included a section for self-reported hours of daily use of the hearing instrument (4, 4–8, 8–12, >12 h). Subjects were required to wear the hearing instrument for at least 6–8 h per day. The number of subjects who reported using their hearing instrument more than 8 h per day was never less than 13 (82%) during any 1-week period.

Next, a listening check was made on the instrument and then electroacoustic tests were carried out to ensure that hearing instrument performance did not change. The tolerance used was 4 dB of the full-on gain (measured with an input level of 50 dB SPL) and OSPL90 measured at the start of the study. In addition, intermodulation distortion was assessed using a composite speech-weighted input signal. None of the hearing instruments tested failed the electroacoustic tests and none was operating with observable intermodulation distortion for the settings and input levels used in the present study. One subject reported a nonfunctioning device between test sessions and a replacement model was fitted the following day. The mean input–output function of the hearing instrument was measured using FAAF-shaped noise. The hearing instrument was positioned at the reference test position in the sound field, and the sound-pressure level was measured in an HA2 2-cc coupler. The output was linear at input levels below 85 dB SPL. As the maximum combined level of speech and noise used in the study was 72 dB SPL, there was approximately 13 dB of headroom and hence hearing instruments were not likely to be operating in significant saturation for any of the speech input levels used.

The first FAAF list of each session was used as a practice run before commencing formal testing. Two FAAF lists (i.e., 160 words) were then used for every test condition: this involved two word lists presented at each level (55, 62, and 69 dB SPL) in the aided and unaided condition for each ear. The subject's own hearing aid was used when measuring aided performance in both the test ear and the control ear. All testing was performed monaurally with the nontest ear plugged and muffed. Benefit scores were obtained by subtracting the unaided score from the aided score. The complete FAAF testing took 100–140 min, including a 20-min break. After the last test session was completed, pure-tone audiometry was repeated to check for any change in hearing sensitivity.

The data were inspected before analysis to confirm it was appropriate to use parametric statistics. Statistical analysis of the mean benefit, aided and unaided scores, consisted primarily of repeated-measures analysis of variance (ANOVA), using SPSS version 10.0, to determine if performance changed significantly over the course of the study. The three factors entered into the ANOVA were ear (test and control), presentation level (55, 62, and 69 dB SPL) and postfitting time (0, 6, and 12 weeks). The degrees of freedom were modified using the Greenhouse–Geisser correction

TABLE V. Mean score (in percent) as a function of postfitting time. One standard deviation is given in parentheses ($n=16$).

Postfitting time (weeks)	0	6	12	Mean
Benefit scores				
Fitted ear				
55 dB SPL	19.0 (8.5)	20.6 (11.0)	21.1 (10.4)	20.2 (9.9)
62 dB SPL	8.0 (7.8)	9.3 (9.8)	11.2 (9.3)	9.5 (8.9)
69 dB SPL	0.0 (7.9)	3.5 (6.2)	5.9 (8.3)	3.1 (7.8)
Mean	9.0 (11.2)	11.1 (11.5)	12.7 (11.1)	10.9 (11.3)
Control ear				
55 dB SPL	17.2 (11.2)	16.8 (11.9)	15.0 (13.0)	16.3 (11.9)
62 dB SPL	7.7 (9.2)	7.3 (10.5)	8.4 (10.1)	7.8 (9.7)
69 dB SPL	1.3 (6.5)	0.7 (11.4)	0.0 (8.2)	0.7 (8.8)
Mean	8.7 (11.1)	8.3 (12.9)	7.8 (12.1)	8.3 (12.0)
Aided scores				
Fitted ear				
55 dB SPL	72.8 (4.5)	74.5 (5.1)	73.8 (5.0)	73.7 (4.9)
62 dB SPL	73.2 (3.8)	75.3 (5.3)	75.8 (4.9)	74.8 (4.7)
69 dB SPL	71.3 (4.9)	73.7 (4.0)	75.0 (3.7)	73.3 (4.4)
Mean	72.5 (4.4)	74.5 (4.8)	74.9 (4.6)	73.9 (4.9)
Control ear				
55 dB SPL	73.0 (6.2)	72.6 (4.8)	70.5 (5.6)	72.0 (5.6)
62 dB SPL	73.9 (6.0)	73.2 (6.0)	71.8 (6.2)	73.0 (6.0)
69 dB SPL	72.7 (5.8)	71.2 (7.7)	70.7 (7.2)	71.5 (6.8)
Mean	73.2 (5.9)	72.3 (6.2)	71.0 (6.2)	72.2 (6.1)
Unaided scores				
Fitted ear				
55 dB SPL	53.8 (7.9)	54.0 (11.7)	52.8 (9.4)	53.5 (9.6)
62 dB SPL	65.3 (6.2)	66.0 (7.6)	64.7 (7.5)	65.3 (7.0)
69 dB SPL	71.3 (4.6)	70.2 (4.7)	69.1 (6.3)	70.2 (5.2)
Mean	63.5 (9.6)	63.4 (10.8)	62.2 (10.3)	63.0 (10.2)
Control ear				
55 dB SPL	56.3 (10.3)	55.8 (10.1)	55.6 (11.6)	55.9 (10.4)
62 dB SPL	66.3 (6.7)	66.0 (7.5)	63.3 (7.5)	65.2 (7.2)
69 dB SPL	71.3 (5.4)	70.5 (7.7)	70.8 (5.7)	70.9 (6.2)
Mean	64.6 (9.9)	64.1 (10.4)	63.2 (10.5)	64.0 (10.2)

when there was a statistically significant deviation from sphericity on Mauchly's test.

III. RESULTS

A. Change in benefit score over time

The mean FAAF benefit (± 1 s.d.), as a function of postfitting time, is given in Table V. The mean benefit scores in the fitted ear with speech presented at 55, 62, and 69 dB SPL at the time of fitting were 19%, 8%, and 0%, respectively. Similar results were obtained in the not-fitted control ear.

The study was concerned primarily with changes in benefit, relative to the time of fitting (Fig. 2). The scores increased for all presentation levels in the fitted ear, with increasing presentation level. By 12 weeks, the mean scores had increased by 2.0%, 3.2%, and 5.9% across the three presentation levels. The corresponding increases in the control ear were all less than 1%.

The mean benefit scores were analyzed using a three-factor (time [3], ear [2], and SPL [3]) repeated-measures ANOVA (see Table VI). As expected, there was a statistically significant difference in mean benefit score with presentation level. The ANOVA orthogonal polynomial breakdown revealed a significant linear component [$F(1,15)=36.7$; $p<0.01$] showing that the slope of the curves in Fig. 2(a)

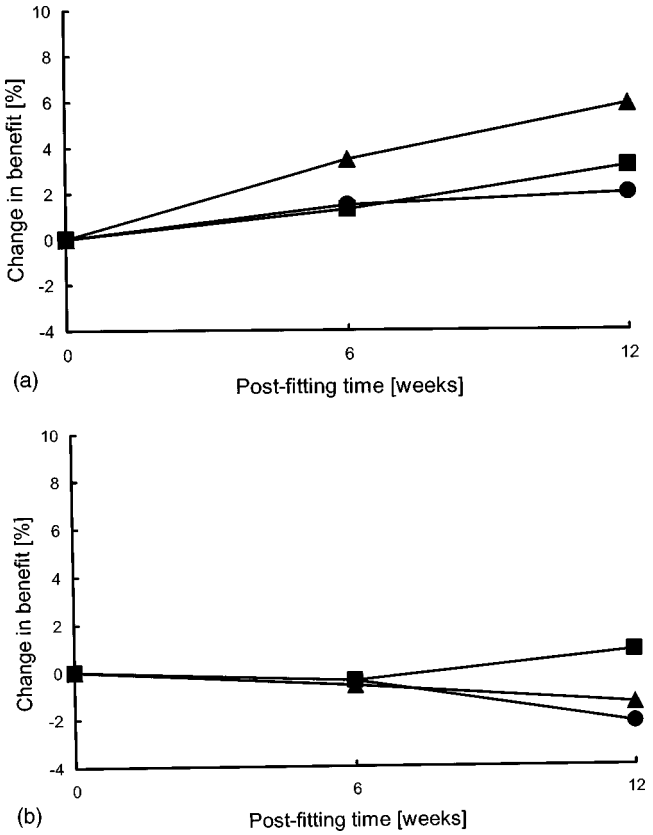


FIG. 2. Mean change in FAAF benefit (aided minus unaided), relative to the time of fitting, as a function of postfitting time with speech at 55 (circles), 62 (squares), and 69 dB SPL (triangles). The fitted ear is shown in (a) and the control ear in (b).

increased with presentation level. The main effects of ear and postfitting time (averaged across other factors) were not statistically significant; however, there was a statistically significant interaction between ear and presentation level. Again, only the linear component of the polynomial breakdown was significant [$F(1,15)=16.2$; $p<0.01$]. This interaction indicates that the fitted and control ears differ over time; the fitted ear showed increasing scores while the control ear showed little change. This was confirmed by investigating each ear separately using a two-factor (time [3] and SPL [3]) repeated measures ANOVA; time was statistically significant in the fitted ear but not the control ear (see Table VII).

Although postfitting time was not significant in the three-factor ANOVA, one-factor (time [3]) repeated-measures ANOVAs performed separately for each presenta-

TABLE VI. Summary of repeated-measures analysis of variance on the benefit scores [$n=16$]. The three within-subject factors were time [3], ear [2], and SPL [3].

Factor	df	F	Significance
Time	2,30	0.9	0.42
Ear	1,15	1.9	0.18
SPL	2,30	31.5	<0.01
Time*Ear	2,30	6.4	<0.01
Time*SPL	4,60	1.1	0.37
Ear*SPL	2,30	0.8	0.46
Time*Ear*SPL	4,60	0.7	0.59

TABLE VII. Summary of simple main effects when the benefit scores from each ear were tested separately [$n=16$]. The two within-subject factors were time [3] and SPL [3].

Factor	df	F	Significance
	Fitted ear		
Time	2,30	4.1	0.03
SPL[a]	1.4,30	32.1	<0.01
Time*SPL	4,60	0.93	0.45
	Control ear		
Time	2,30	0.35	0.71
SPL[a]	1.3, 20.2	21.9	<0.01
Time*SPL	4,60	0.84	0.50

[a]Mauchly test significant; degrees of freedom adjusted using Greenhouse–Geisser correction.

tion level showed no statistically significant change in benefit at 55 dB [$F(2,30)=0.5$, $p>0.05$], but there was a statistically significant increase in benefit at 69 dB [$F(2,30)=5.6$, $p<0.01$]. The improvement in benefit in the fitted ear over time was not simply due to skewing by a small number of subjects; the change in benefit at the highest presentation level ranged from −4% to +19%, with eight (50%) subjects showing an increase in excess of the mean improvement of +5.9%.

B. Aided performance over time

The mean aided and unaided scores (±1 s.d.) from which the benefit scores are calculated are also given in Table V. The initial aided scores were around 72% in both ears. The changes in score are shown in Fig. 3 for the fitted ear and the control ear. The aided scores increased over time in the fitted ear for all presentation levels (by 1.0%, 2.6%, and 3.6% at 55, 62, and 69 dB SPL, respectively). Changes in the control ear were all negative and typically around −2.0%.

The mean aided scores were analyzed using a three-factor (time [3], ear [2], and SPL [3]) repeated-measures ANOVA (see Table VIII). The main effect of ear was statistically significant; the mean score (averaged across other factors) in the fitted ear was 73.9% compared with 72.2% in the control ear. The linear component of the ANOVA polynomial breakdown was significant [$F(1,15)=6.0$; $p=0.03$]. In addition, there was a statistically significant interaction between ear and postfitting time [$F(2.30)=15.5$; $p<0.01$]. This interaction indicates that the fitted and control ears differ over time; the fitted ear showed increasing scores while the control ear showed slightly decreasing scores. Again, only the linear component of the polynomial breakdown was significant [$F(1,15)=27.0$; $p<0.01$]. Two-factor (time [3] and SPL [3]) repeated-measures ANOVA performed on each ear separately (see Table IX) revealed postfitting time to be statistically significant for each ear, but there was no interaction with presentation level. The linear component of the polynomial breakdown was significant for both ears [fitted ear, $F(1,15)=13.4$; $p<0.01$; control ear, $F(1,15)=7.4$, $p=0.02$].

Consistent with the benefit data, when ANOVAs were performed separately for each, presentation level showed no

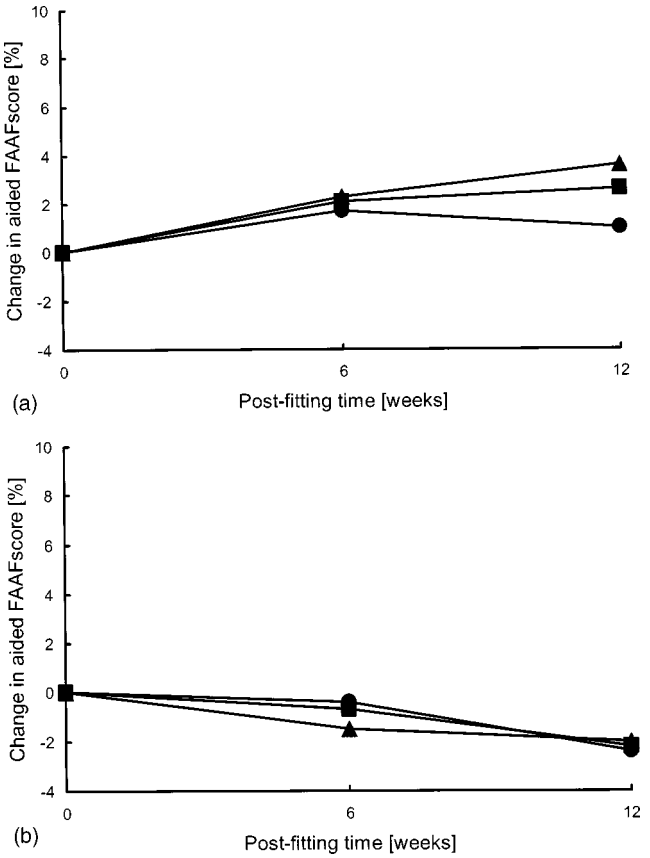


FIG. 3. Mean change in aided FAAF score (%) relative to fitting, at a speech presentation level of 55 (circles), 62 (squares), and 69 dB SPL (triangles). The fitted ear is shown in (a) and the control ear in (b).

statistically significant change at 55 dB [$F(2,30)=1.2$,$p>0.05$] but there was a statistically significant increase in benefit at 69 dB [$F(2,30)=6.5$,$p<0.01$]. The improvement in aided performance in the fitted ear was not simply due to skewing by a small number of subjects; the change in aided performance at the highest presentation level ranged from −4% to +14% with nine (56%) subjects showing an increase in excess of the mean improvement of +3.6%.

C. Unaided performance over time

The mean unaided scores (±1 s.d.), as a function of postfitting time, are also given in Table V. The initial performance was typically around 55% with speech at 55 dB SPL but nearer 71% with speech at 69 dB SPL. The change in

TABLE VIII. Summary of repeated-measures analysis of variance on the aided scores [$n=16$]. The three within-subject factors were time [3], ear [2], SPL [3].

Factor	df	F	Significance
Time	2,30	0.6	0.55
Ear	1,15	6.0	0.03
SPL[a]	1.3,19.5	1.1	0.33
Time*Ear	2,30	15.5	<0.01
Time*SPL	4,60	1.3	0.30
Ear*SPL	2,30	0.0	0.99
Time*Ear*SPL	4,60	0.4	0.80

[a]Mauchly test significant; degrees of freedom adjusted using Greenhouse–Geisser correction.

TABLE IX. Summary of simple main effects on the aided scores when each ear was tested separately [$n=16$]. The two within-subject factors were time [3] and SPL [3].

Factor	df	F	Significance
	Test ear		
Time	2,30	7.4	<0.01
SPL[a]	1.4,21.2	1.0	0.36
Time*SPL	4,60	1.1	0.38
	Control ear		
Time	2,30	4.9	0.02
SPL[a]	1.3,30	0.8	0.43
Time*SPL	4,60	0.5	0.75

[a]Mauchly test significant; degrees of freedom adjusted using Greenhouse–Geisser correction.

score relative to the time of fitting is shown in Fig. 4 for both ears. There was a trend for scores to decrease slightly over time for all presentation levels in both ears, but this was not significant (see Table X).

IV. DISCUSSION

A. Speech recognition performance at the time of fitting

The magnitudes of the derived benefit scores (aided minus unaided) are greater for the lower presentation levels. The unaided scores are greater at higher presentation levels whereas the aided scores are similar across levels; this has

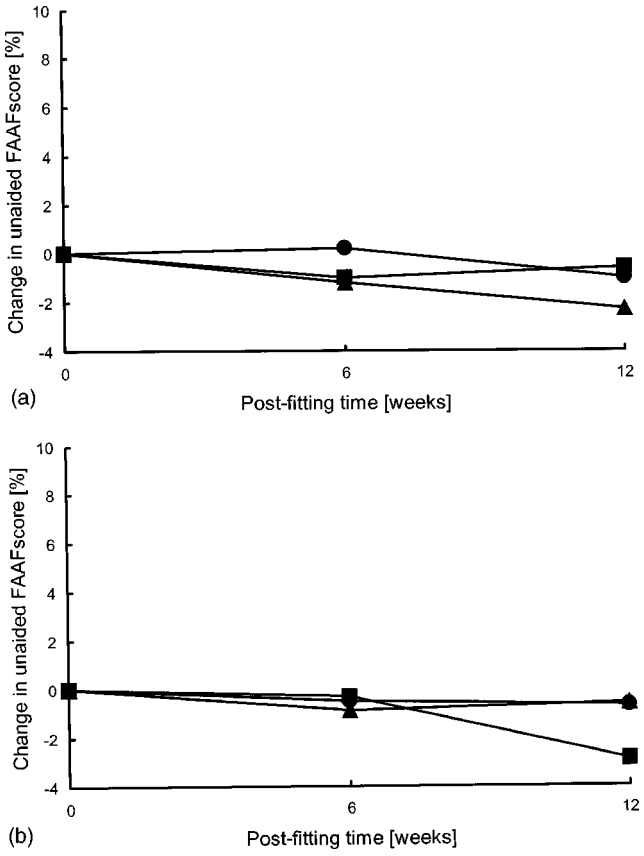


FIG. 4. Mean change in unaided FAAF score (%) relative to fitting, at a speech presentation level of 55 (circles), 62 (squares), and 69 dB SPL (triangles). The fitted ear is shown in (a) and the control ear in (b).

TABLE X. Summary of repeated-measures analysis of variance on the unaided scores [$n=16$]. The three within-subject factors were time [3], ear [2], and SPL [3].

Factor	df	F	Significance
Time	2,30	1.4	0.26
Ear	1,15	0.4	0.56
SPL	2,30	42.0	<0.01
Time*Ear	2,30	0.1	0.89
Time*SPL	4,60	0.6	0.67
Ear*SPL	2,30	0.8	0.45
Time*Ear*SPL	4,60	0.6	0.66

the effect of reduced benefit score at the higher level. This is expected since performance in the unaided condition is mediated mainly by audibility, whereas in the aided condition most speech is audible and performance is limited equally for all levels by the competing FAAF-shaped noise. The mean benefit scores at the time of fitting were comparable in both ears, consistent with the symmetrical hearing loss.

Mean aided performance with speech presented at 69 dB SPL was 1%–2% lower than at the two lower presentation levels. The decrement was small and not statistically significant. Some studies indicate that performance remains constant at high speech presentation levels (for example, Duquesnoy and Plomp, 1983), while others indicate that performance is reduced (for example, Larson *et al.*, 2000; Studebaker *et al.*, 1999). This latter ''roll-over'' effect is presumably related to increased upward spread of masking resulting in vowels masking consonants.

B. Change in benefit over time

Following 12 weeks of hearing instrument use there was a systematic increase in benefit of around 4% (averaged across presentation level) in the fitted ear. There was no change in the control ear. This means that the increase in benefit cannot be explained in terms of simple practice effects. The significant increase in benefit over time replicates the findings of studies discussed earlier that demonstrate acclimatization.

The only previous study that used the not-fitted ear as a control was Gatehouse (1992); this also showed no increase in the control condition. This finding is consistent with the specific nature of perceptual learning discussed by Robinson and Summerfield (1996). Studies where human subjects are trained on frequency discrimination (Demany, 1985; Irvine *et al.*, 2000) temporal discrimination (Wright *et al.*, 1997), and sound localization (Wright and Fitzgerald, 2001) tasks show there is some degree of specificity to the training stimulus. This suggests that any learning associated with the provision of a monaural hearing instrument may be specific to the fitted ear and may not transfer to the control ear.

The magnitude of the acclimatization effect observed in the present study is consistent with most of the previous studies. The one exception is the study by Gatehouse (1992) that reported a 14% increase in benefit at 12 weeks postfitting. Gatehouse reported a mean unaided score of around 71% and a benefit score of around 4% at the time of fitting.

Similar scores were obtained in the present study with a presentation level of 69 dB SPL. At this presentation level, the change in mean benefit in the present study was 5.9%. A possible explanation for the apparent discrepancy is the difference in the REIG provided in the two studies; Gatehouse provided approximately 12 dB more gain than the present study. This means that there was a larger difference in level between the amplified and unamplified speech. This may mean that more reorganization was required by the subjects in the Gatehouse study, resulting in greater improvement over time.

The implications for clinical practice are unclear. Several authors (for example, Gatehouse, 1997, 1998; Humes *et al.*, 2001) have demonstrated that word recognition scores, in isolation, give an incomplete representation of hearing aid benefit. In addition, the increase in mean aided performance for average speech was only +2.6%. The study cannot speak to longer time scales because there is no evidence of performance reaching a plateau by 12 weeks postfitting. This finding is consistent with other studies conducted over a similar time scale (Gatehouse, 1992; Horwitz and Turner, 1997).

C. Changes in aided and unaided scores over time

The change in benefit score in the fitted ear was due to both an increase in aided performance and a decrease in unaided performance, although only the former reached statistical significance. There are mixed findings in the acclimatization literature regarding changes in the unaided condition. The studies by Cox *et al.* (1996) and Horwitz and Turner (1997) do not show any reduction in the unaided scores over a 12–18-week postfitting period (although a subgroup of subjects from the Cox *et al.* study was reviewed 9 months postfitting and the results showed a dramatic reduction in unaided performance). The clearest demonstration of a reduction in performance in the unaided condition (approximately 8%) comes from Gatehouse (1992). This discrepancy may also be related to the greater amount of REIG provided by Gatehouse. If the fitted ear becomes accustomed to hearing speech at a high amplified level, then over time it will perform worse at the less familiar level of unaided speech.

In addition to the changes observed in the fitted ear, there was a progressive decrease in performance over time in the control ear. By 12 weeks, the mean aided scores showed a statistically significant decrease of 2.2%, while the unaided scores showed a trend towards reduced performance of 1.4%. This suggests that the control ear was deprived of adequate stimulation relative to the fitted ear. This finding is consistent with retrospective reports of late onset auditory deprivation (see the review by Neuman, 1996), although it occurred over a much shorter time scale in the present study. However, a similar finding was reported by Gatehouse (1992), who showed a statistically significant reduction in performance in three out of six test conditions in the control ear over a postfitting time period of 12 weeks. It may be that resources within the auditory system are diverted to subserve the fitted ear at the expense of the other ear.

D. The influence of presentation level

The results show greater increases in benefit over time at the higher presentation levels. This is supported by the one-factor (time) repeated-measures ANOVAs performed separately for each presentation level. The lack of statistically significant interaction between presentation level and time occurred because the statistical analysis was underpowered. The mean difference between the increase in benefit at 55 and 69 dB SPL was 3.8% but the standard deviation was 8.3%. The higher than expected variability means that the statistical power was only around 40%: this resulted in a material risk of a type II error. Approximately double the number of subjects would be required to increase the statistical power to 80%. Alternatively, the effect size may have increased if the design had been modified to include a longer postfitting time period, a wider range of presentation levels, or if subjects were provided with greater REIG.

A summary of an experiment that measured the acclimatization effects with linear and nonlinear signal processing was included in an article by Gatehouse (1998). One finding was that the acclimatization effect with linear hearing instruments differed with presentation level: acclimatization was shown at a presentation level of 65 and 75 dB SPL but not at 55 dB SPL. While the study has yet to be published in detail, this finding is consistent with the level-dependent effects observed in the present experiment.

Several previous studies have failed to show an acclimatization effect despite using a presentation level similar to the highest level used in the present study. Bentler *et al.* (1993) used a speech presentation level between 50–60 dB HL. When transformed from hearing level to sound-pressure level, the presentation level is approximately 65–70 dB. However, there are potentially important methodological differences between the two studies that could explain the difference in outcome. For example, Bentler *et al.* included subjects with a relatively mild hearing impairment, some were existing hearing aid users, the frequency response failed to meet NAL-RM targets above 2 kHz, and the gain control was not fixed across time. Despite these differences, and the lack of statistically significant improvements in mean performance over time, several of the performance measures (low predictability SPIN sentences and NST in noise) showed an increase of around 4% at 1–3 months postfitting; this is comparable with the present study.

There are a number of additional studies that demonstrate a level-dependent effect after hearing instrument use. Gatehouse (1989) measured the SNR required for an unaided performance of 50% on FAAF in 24 monaurally aided subjects and showed an interaction between presentation level (65–90 dB SPL) and ear. Criterion performance was obtained with a less favorable SNR when the fitted ear was tested at high levels compared to the not-fitted ear, while the reverse occurred at lower presentation levels. This intensity dependence was taken as evidence that an ear performs most efficiently at its familiar listening level. The difference in SNR between ears at the highest presentation level of 90 dB SPL was 2.9 dB. The difference in aided score between the two ears in the present study, at a presentation level of 69 dB SPL, was 5.6% (test ear +3.6%, control ear −2.0%). To

allow a comparison with the change in SNR from the Gatehouse study, these differences in score can be converted into differences in SNR using the average FAAF performance intensity function obtained from hearing-impaired subjects as reported by Shields and Campbell (2001). At a performance level of 70%, a change of 3% occurs for every 1-dB change in SNR. Therefore, the difference between ears of 5.6% corresponds to a difference in SNR ears of approximately 2 dB. Thus, the magnitude of change expressed in terms of SNR is similar to that reported by Gatehouse.

In 1995, Robinson and Gatehouse (1995) reported the results of a retrospective study that showed that the fitted ear of subjects aided monaurally was better able to discriminate intensity than the not-fitted ear at levels above 95 dB SPL; in contrast, the not-fitted ear was better able to discriminate intensity than the fitted ear at lower levels. In 1996, Robinson and Gatehouse (1996) reported a prospective study that measured the ability to discriminate changes in intensity in five new users of monaural hearing instruments. After 12–18 weeks of hearing instrument use, the subjects displayed level- and frequency-dependent changes in intensity discrimination for the fitted ear but not the control ear. Specifically, the fitted ear was shown to perform better at high levels but worse at lower levels. These changes only occurred at the frequencies that received material benefit from amplification.

Gatehouse and Robinson (1996) subsequently reported an experiment that showed changes in loudness function in four subjects who were fitted monaurally. Specifically, sounds were rated louder in the fitted ear than in the control ear over the range of intensities studied. Again, the changes were confined to the high frequencies where the hearing instrument provided benefit. In the same study, the authors also reported a single case study where the changes in intensity discrimination were accompanied by changes in amplitude of the N1-P2 complex of the slow vertex cortical-evoked response. At a high intensity level the amplitude was greater in the fitted ear but smaller at lower intensity levels. Taken collectively, these studies show a level-dependent effect and this has generally been explained in terms of the ear performing most efficiently at the presentation level to which it has become accustomed.

Before fitting, the ears of the subjects used in the present study would be expected to be performing most efficiently over the range 55 to 69 dB SPL since this is assumed to cover the range of levels experienced in real life. After fitting, this range will be shifted to a higher level at the eardrum; however, there is an overlap between the range of levels experienced when aided and unaided (raised speech before aiding is similar in level to quiet speech after aiding). Hence, only the level of aided speech at 69 dB SPL is ''new'' to the subjects and it is for this reason that a period of acclimatization is required for this condition.

An alternative explanation for the maximum acclimatization effect occurring at the highest presentation level is related to the subject's ability to extract speech from background noise. Since the SNR was fixed across presentation levels, the level of noise increased along with the speech presentation level. Thus, the maximum acclimatization effect occurred when listening in the presence of the highest noise levels. It is possible that the subject required time to become able to extract the speech signal from the high-level background noise. For this to be correct, the subject would have to be exposed to high levels of noise in every day life. The FAAF-shaped noise was used in the study to limit the speech recognition scores at the time of fitting: it was not selected to be representative of everyday listening conditions. Although the noise levels experienced by the subjects in their daily environments were not measured, Pearsons *et al.* (1977) reported the SNRs measured for conversational speech in and around urban homes to be of the order of $+5$ to $+9$ dB. This is more favorable than the median of -2 dB used in the present study. Thus, it appears unlikely that there would have been sufficient opportunity (in every day life) for subjects to learn to extract a signal at the negative SNRs used in the present study. In addition, this does not explain why performance decreased over time in the fitted ear when the hearing instrument was removed. It should be possible to design an experiment that involves speech material being presented with and without masking noise (but avoiding ceiling effects). If performance were to increase under both conditions, then acclimatization cannot be due solely to the ability to extract speech from noise.

It is interesting to speculate on the physiological mechanism responsible for acclimatization. It is possible that acclimatization may entail reallocation of resources to areas of the auditory system representing high-level stimuli. This is consistent with the progressive improvements in intensity discrimination reported by Gatehouse and Robinson (1996).

With the onset of a hearing impairment, one might envisage that high-threshold regions within the auditory system may become underused because the sound levels seldom reach sufficient levels to activate them. This might lead to the lower threshold regions invading the underused areas due to lack of competition from high stimulus levels. After provision of amplification, the auditory system has to reorganize or reacclimatize to the range of high signal levels. The effect may be that the auditory system increases its representation of high levels of speech, at the expense of the previously lower unaided levels, in order to make maximum use of the newly amplified speech signal. This suggests that larger changes in listening level will result in greater reorganization. This may explain why Gatehouse reported a larger acclimatization effect in subjects who were fitted with hearing instruments that provided greater gain than in other studies. Reorganization in the intensity domain is consistent with the remapping of the relationship between intensity and loudness reported by Gatehouse and Robinson (1996); the greatest capacity for discriminating differences in intensity shifts from lower to higher intensities as subjects gain amplification experience.

In summary, it is hypothesized that, after provision of a hearing instrument, the auditory system reallocates its resources in the intensity domain, in addition to changes in the frequency domain. This results in an increased representation of the behaviorally important speech sounds that are now presented at a higher intensity than previously with a corresponding reduction in representation of lower intensities.

These processes take time to occur and can explain the phenomenon of auditory acclimatization. Because the reallocation in the intensity domain involves the upper end of the intensity range, auditory acclimatization is most readily demonstrated using speech at relatively high presentation levels. The present study demonstrates clear acclimatization for speech at 69 dB but minimal acclimatization for speech at 55 dB, consistent with the above explanation.

V. CONCLUSIONS

The results of this study confirm the presence of auditory acclimatization predominantly when using speech in noise at a high presentation level. This suggests that studies of auditory acclimatization must use a level of amplified speech signal that is higher than commonly experienced in everyday life, prior to aiding. Further experiments are desirable to confirm this effect across a wider range of presentation levels.

ACKNOWLEDGMENTS

This work was supported by a research studentship from the NHS Research and Development Directorate in the UK.

ANSI (**1989**). ANSI S3.6 (1989), ''Specification for Audiometers'' (American National Standards Institute, Inc., New York).

ANSI (**1997**). ANSI S3.5 (1997), ''American National Standard Methods for the Calculation of the Speech Intelligibility Index'' (American National Standards Institute, Inc., New York).

Arlinger, S., Gatehouse, S., Bentler, R. A., Byrne, D., Cox, R. M., Dirks, D. D., Humes, L., Neuman, A., Ponton, K., Robinson, K., Silman, S., Summerfield, A. Q., Turner, C. W., Tyler, R. S., and Willott, J. F. (**1996**). ''Report of the Eriksholm workshop on auditory deprivation and acclimatization,'' Ear Hear. **17**, 87S–98S.

Bentler, R. A., Neibuhr, D. P., Getta, J. P., and Anderson, C. V. (**1993**). ''Longitudinal study of hearing aid effectiveness. I. Objective measures,'' J. Speech Hear. Res. **36**, 808–819.

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.

Cox, R. M., and Alexander, G. C. (**1992**). ''Maturation of hearing aid benefit: Objective and subjective measurements,'' Ear Hear. **13**, 1311–1341.

Cox, R. M., Alexander, G. C., Taylor, I. M., and Gray, G. A. (**1996**). ''Benefit acclimatization in elderly hearing aid users,'' J. Am. Acad. Audiol **7**, 428–441.

Demany, L. (**1985**). ''Perceptual learning in frequency discrimination,'' J. Acoust. Soc. Am. **78**, 1118–1120.

Duquesnoy, A. J., and Plomp, R. (**1983**). ''The effect of a hearing aid on the speech reception threshold of hearing-impaired listeners in quiet and in noise,'' J. Acoust. Soc. Am. **73**, 2166–2173.

Foster, J. R., and Haggard, M. P. (**1979**). ''FAAF. An effective analytical test of speech perception,'' Proc. Inst. Acous. **182**, 9–12.

Foster, J. R., and Haggard, M. P. (**1987**). ''The four alternative auditory feature test (FAAF): Linguistic and psychometric properties of the material with normative data in noise,'' Br. J. Audiol. **21**, 165–174.

Gatehouse, S. (**1989**). ''Apparent auditory deprivation effects of late onset: The role of presentation level,'' J. Acoust. Soc. Am. **86**, 2103–2106.

Gatehouse, S. (**1992**). ''The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids,'' J. Acoust. Soc. Am. **92**, 1258–1268.

Gatehouse, S. (**1993**). ''Role of perceptual acclimatization in the selection of frequency responses for hearing aids,'' J. Am. Acad. Audiol **4**, 296–306.

Gatehouse, S. (**1997**). ''Outcome measures for the evaluation of adult hearing aid fittings and services,'' Scientific and Technical Report to the Department of Health, London, England.

Gatehouse, S. (**1998**). ''Speech tests as measures of outcome,'' Scand. Audiol. Suppl. **27**, 54–60.

Gatehouse, S., and Robinson, K. (**1996**). ''Acclimatization to monaural hearing aid fitting effects on loudness functions and preliminary evidence for parallel electrophysiological and behavioural effects,'' in *Psychoacoustics, Speech and Hearing aids*, edited by B. Kollmeier (World Scientific, Singapore), pp. 319–330.

Hawkins, D., Alvarez, E., and Houlihan, J. (**1991**). ''Reliability of three types of probe tube microphone measurements,'' Hear. Instru. **42**, 14–16.

Horwitz, A. R., and Turner, C. W. (**1997**). ''The time course of hearing aid benefit,'' Ear Hear. **18**, 1–11.

Humes, L. E., Barlow, N. N., Garner, C. B., and Wilson, D. L. (**2000**). ''Prescribed clinician-fit versus as-worn coupler gain in a group of elderly hearing-aid users,'' J. Speech Lang. Hear. Res. **43**, 879–892.

Humes, L. E., Garner, C. B., Wilson, D. L., and Barlow, N. N. (**2001**). ''Hearing-aid outcome measures following one month of hearing aid use by the elderly,'' J. Speech Lang. Hear. Res. **44**, 469–486.

Humes, L. E., Halling, D., and Coughlin, M. (**1996**). ''Reliability and stability of various hearing-aid outcome measures in a group of elderly hearing-aid wearers,'' J. Speech Hear. Res. **39**, 923–935.

Humes, L. E., Wilson, D. L., Barlow, N. N., and Garner, C. (**2002**). ''Changes in hearing-aid benefit following one, two, or three years of hearing aid use by the elderly,'' J. Speech Hear. Res. (in press).

ISO 8253-2. (**1998**). *Acoustics—Audiometric test methods. Part 2: Sound Field Audiometry with Pure Tone and Narrow-band Test Signals* (International Organization for Standardization, Geneva, Switzerland).

Irvine, D. R. F., Martin, R. L., Klimkeit, E., and Smith, R. (**2000**). ''Specificity of perceptual learning in a frequency discrimination task,'' J. Acoust. Soc. Am. **108**, 2964–2968.

Larson, V. D., Williams, D. W., Henderson, W. G., Leuthke, L. E., Beck, L. B., Noffinger, D., Wilson, R. H., Dobie, R. A., Haskell, G. B., Bratt, G. W., Shanks, J. E., Stemmoachowicz, P., Studebaker, G. A., Boysen, A. E., Donahue, A., Canalis, R., Fausti, S. A., and Rappaport, B. Z. (**2000**). ''Efficacy of three commonly used hearing aid circuits,'' J. Am. Med. Assoc. **11**, 1806–1813.

Lutman, M. E., and Clark, J. (**1986**). ''Speech identification under simulated hearing-aid frequency response characteristics in relation to sensitivity, frequency resolution, and temporal resolution,'' J. Acoust. Soc. Am. **80**, 1030–1040.

Munro, K. J., and Lutman, M. E. (**2000**). ''Auditory acclimatization to amplified speech,'' Poster presentation at IHCON 2000, Lake Tahoe.

Neuman, A. C. (**1996**). ''Late-onset auditory deprivation: A review of past research and an assessment of future research needs,'' Ear Hear. **17**, 3S–13S.

Palmer, C. V., Nelson, C. T., and Lindley, G. A. (**1998**). ''The functionally and physiologically plastic adult auditory system,'' J. Acoust. Soc. Am. **103**, 1705–1721.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). ''An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,'' J. Acoust. Soc. Am. **80**, 50–57.

Pearsons, K. S., Bennett, R. L., and Fidells, S. (**1977**). ''Speech levels in various noise environments'' (Report No. EPA-600/1-77-025). (U.S. Environmental Protection Agency, Washington, D.C.).

Robinson, D. W. (**1991**). ''Long-term repeatability of the pure-tone hearing threshold and its relation to noise exposure,'' Br. J. Audiol. **25**, 219–235.

Robinson, K., and Gatehouse, S. (**1995**). ''Changes in intensity discrimination following monaural long-term use of a hearing aid,'' J. Acoust. Soc. Am. **97**, 1183–1190.

Robinson, K., and Gatehouse, S. (**1996**). ''The time course of effects on intensity discrimination following monaural fitting of hearing aids,'' J. Acoust. Soc. Am. **99**, 1255–1258.

Robinson, K., and Summerfield, A. Q. (**1996**). ''Adult auditory learning and training,'' Ear Hear. **17**, 51S–65S.

Saunders, G. H., and Cienkowski, K. M. (**1997**). ''Acclimatization to hearing aids,'' Ear Hear. **18**, 129–139.

Shields, P. W., and Campbell, D. R. (**2001**). ''Intelligibility, subjective ratings and completion time scores using the FAAF test with hearing-impaired subjects and noisy reverberant environments,'' Br. J. Audiol. **35**, 237–245.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). ''Monosyllabic word recognition at higher-than-normal speech and noise levels,'' J. Acoust. Soc. Am. **105**, 2431–2444.

Surr, R. K., Cord, M. T., and Walden, B. E. (**1998**). ''Long-term versus short-term hearing aid benefit,'' J. Am. Acad. Audiol. **9**, 165–171.

Taylor, K. S. (**1993**). ''Self-perceived and audiometric evaluations of hearing aid benefit in the elderly,'' Ear Hear. **14**, 390–394.

Turner, C. W., and Bentler, R. A. (**1998**). ''Does hearing aid benefit increase over time?'' Letter to Editor. J. Acoust. Soc. Am. **104**, 3673–3674.

Turner, C. W., Humes, L. E., Bentler, R. A., and Cox, R. M. (**1996**). ''A review of past research on changes in hearing aid benefit over time,'' Ear Hear. **17**, 14S–28S.

Wright, B. A., Buonomano, D. V., Mahnvcke, H. W., and Merzenich, M. M. (**1997**). ''Learning and generalization of auditory temporal-interval discrimination in humans,'' J. Neurosci. **17**, 3956–3963.

Wright, B. A., and Fitzgerald, M. B. (**2001**). ''Different patterns of human discrimination learning for two interaural cues to sound-source localization,'' Proc. Nat. Acad. Sci. U.S.A. **98**, 12307–12312.

Delayed models for simplified musical instruments

Ana Barjau[a)]
Departament d'Enginyeria Mecànica, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

Vincent Gibiat
LAMI, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex, France

(Received 28 May 2002; revised 20 March 2003; accepted 7 April 2003)

Most musical instruments contain, at their very basis, a continuous vibrating element (string or air column) which can be treated as a one-dimensional system. Its oscillation is obtained either through an initial condition or by means of a continuous energy input through a nonlinear device. In both cases and as a first approach, the excitation can be localized at one single point, and the continuous system can be considered as a linear one. The coupling between these two elements is often represented through a convolution integral. This convolution will be rewritten here in a way that different phenomena taking place in the continuous element (internal losses, radiation at the ends...) are separated. Different choices in the formulation of these processes and some mathematical manipulation will lead to either algebraic iterative or delayed differential equations. These equations are valid for any form of energy input. Once this energy input is defined, they can be used to simulate the behavior of different instruments in a more efficient way than that of traditional convolution. Moreover, these equations allow an analytical analysis of possible regimes using the tools of nonlinear dynamical systems (NLDS). The case of woodwinds will be emphasized throughout the paper, while that of strings will be presented briefly for the sake of completeness.
 [DOI: 10.1121/1.1577558]

PACS numbers: 43.75.De, 43.75.Ef, 43.25.Ts [NHF]

I. INTRODUCTION

The general structure of most musical instruments is a continuous resonant system (air column or string) coupled to an energy source (the player) and sometimes to a secondary resonant system (for the case of the string family). The most outstanding difference between the self-sustained instruments and those based on free oscillations is the essential nonlinearity existing in the former case whose role is to transform the constant energy flow into an oscillating one. As a first approach, both the initial conditions (for the case of instruments based on free oscillations) and the nonlinearity (for self-sustained instruments) can be localized at one single point (the excitation point), and the continuous system can be considered as a linear one. The coupling between these two elements leads to different oscillating regimes.

The aim of this paper is to establish a general time-domain formulation for drastically simplified wind and string instruments. As a basic tool, this formulation will allow analytical predictions difficult to obtain in the frequency domain. It is clear that those results will be of no help in synthesis or instruments design, but as a conceptual tool they should provide some hints to understand the different behaviors possible in those ''simple instruments.'' For the case of real self-sustained instruments, those behaviors are rich and complex, and so preliminary work on simplified models is necessary before studying more realistic structures. Moreover, as the mathematical formulations we will establish are general, our results could probably be extrapolated to other fields of physics. The idealization for the wind family description will lead to purely cylindrical or purely conical bores without holes, while that for the string family will consist in ignoring the existence of any resonant part other than the string itself. The dissipation and radiation processes will be represented mathematically through convolutions with exponential kernels. Of course, those models can be eventually used for efficient numerical simulations, but their simplicity will not allow questions of direct significance to performers and instrument makers to be addressed.

The main point of this work is the treatment of the continuous linear system. Thus, for the case of self-sustained instruments, the formulation of the nonlinearity will not be discussed. We will just assume that there is a nonlinear relationship between the variable $g(t)$ (the output of the nonlinear device) and the variable $f(t)$ that will be taken as the response of the continuous system at the excitation point, $g(t)=NL[f(t)]$. A second paper will deal with the self-sustained regimes that can be obtained from those models and different choices of the nonlinear characteristics.

The linear element is, strictly speaking, a 3D continuous system. However, the usual models that can be found in the scientific literature are 1D simplifications. Some authors (Antunes *et al.*, 2000; Agulló and Barjau, 1986; Barjau and Gibiat, 1997b) work with a modal analysis approach (and so a finite number of degrees of freedom—DOF) while others retain the continuous essence of the system (and so work with an infinite number of DOFs) (Martínez and Agulló, 1988; Schumacher, 1981). Our approach will also be a 1D continuous one.

As we have chosen a localized description for the exci-

a)Electronic mail: Ana.Barjau@upc.es

tation, it seems logical to adopt the convolution as the more suitable description of the coupling between $g(t)$ and $f(t)$. The traditional form of this equation is

$$f(t)=h(t)*g(t), \tag{1}$$

where $h(t)$ is the impulse response of the system.

Equation (1) has two disadvantages. On one hand, as the kernel has a low decaying rate in most cases, the integration calls for the whole past history of $g(t)$ and so is capacity and time consuming. On the other hand, all the different physical processes related to losses (internal and radiation losses, energy transfer to other systems) and propagation are mixed up in the expression of $h(t)$.

The first point has already been addressed by several authors (Schumacher, 1981; Agulló *et al.*, 1988; McIntyre and Woodhouse, 1979). We will propose a formulation based on ordinary differential equations (ODE). It is clear that such approach does not have this drawback: once the integration method has been chosen, the number of operations per time step remains constant.

Concerning the second point, we will work from the beginning with separate formulations for each physical phenomenon related to propagation and loss mechanisms, and so the final convolution equation will allow us to work with different and progressive approaches (models with just radiation losses, models with radiation plus internal losses, models with or without dispersion...). More precisely, any change in the shape of the propagating functions associated with internal losses in general will be represented through a kernel function $\sigma_i(t-L/c)$ (where c is the wave propagation speed) which will depend, among other features, on the length L of the 1D system.

Whenever a local characteristic of the system is responsible for an immediate spreading of the input signal at the excitation point [such as taper in conical bores or bending stiffness and torsional modes in strings (Woodhouse and Loach, 1999)], a local kernel function $\theta(t)$ will be used that will take into account such effect.

Finally, the energy transfer at the open ends (for the case of wind instruments) or through the supports (for the string family) will be taken into account through local elementary reflection functions $R_i(t)$.

As a first step, we will rewrite the traditional convolution using these general kernel functions. We will show that if the kernels are expressed as a combination of Dirac delta distributions and exponential functions, a Laplace transform of this basic general integral equation leads to either an iterative algebraic equation or a delayed differential equation (that is, an ODE containing delayed variables).

Some of the equations that will be obtained have already been published in former works. Maganza *et al.* (1986) presented a non dispersive iterative algebraic model for cylindrical wind instruments. Friedlandler (1953) did the same for bowed strings. We will reobtain those results as a particular case of our general formulation. Barjau *et al.* (1997a) obtained two delayed differential equations for cylindrical and conical woodwinds with just radiation losses, which will also be presented here as particular cases.

II. THE GENERAL CONVOLUTION

The main difference between wind and stringed instruments is that in the former case the excitation point is generally placed at one end of the linear resonant system (the air column) while in the latter it can be placed anywhere between the two string ends. The consequence is that the behavior of a wind instrument will depend on one characteristic time τ (the time needed for a round trip along the bore, $\tau=2L_0/c$, where L_0 is the total geometrical length of the system and c is the propagation speed), while that of a string instrument will rely on two characteristic times τ_1 and τ_2 (those corresponding to the round trips between the excitation point and the two ends of the string: $\tau_i=2L_i/c$, where L_i is the length of each string portion).

It is worth pointing out that an accurate model of flute would also take into account two time delays (Gibiat, 1988).

A. Wind instruments description

For the case of wind instruments, the input and output variables $g(t)$ and $f(t)$ are coupled through the classical convolution equation (1). The impulse response of the system $h(t)$ can be understood as the whole time evolution of $f(t)$ due to a $g(t)$ impulse.

From a physical point of view the behavior of a conical bore is totally different from that of a cylindrical one. Regardless of the intensity of internal losses, the $f(t)$ response at the input section of a divergent cone always shows an exponential decaying wake that does not appear in the case of a lossless cylinder. In order to treat both geometries through the same equation, it is useful to separate the impulse response into an ''early'' response and a ''long-term'' response

$$h(t)=h(t<\tau)+h(t\geqslant\tau)\equiv h_0(t)+h_+(t). \tag{2}$$

Independently from the particular form of $h_0(t)$, the long-time response $h_+(t)$ can be constructed through a propagation–reflection rationale. If we call $R_1(t)$ and $R_2(t)$ the elementary reflection functions associated with the input and output ends, respectively, and we take into account the fact that each propagation along the bore implies a time delay $\tau=2L_0/c$ and a dissipation [represented through the kernel function $\sigma_0(t)$], then $h_+(t)$ can be written as

$$h_+(t)=\sum_{n=1}^{\infty}[R_1(t)*]^{n-1}[R_2(t)*]^n[\sigma_0(t)*]^n h_0(t-n\tau) + \sum_{n=1}^{\infty}[R_1(t)*]^n[R_2(t)*]^n[\sigma_0(t)*]^n h_0(t-n\tau). \tag{3}$$

The first summation corresponds to the upward wave (that propagating from the output to the input section), whereas the second one corresponds to the downward wave (propagating from the input to the output section).

Substitution of (2) and (3) into (1) gives

$$f(t)=f_0(t)+[\delta(t)+R_1(t)]*\sum_{n=1}^{\infty}[R_1(t)*]^{n-1}\times[R_2(t)*]^n[\sigma_0(t)*]^n f_0(t-n\tau), \tag{4}$$

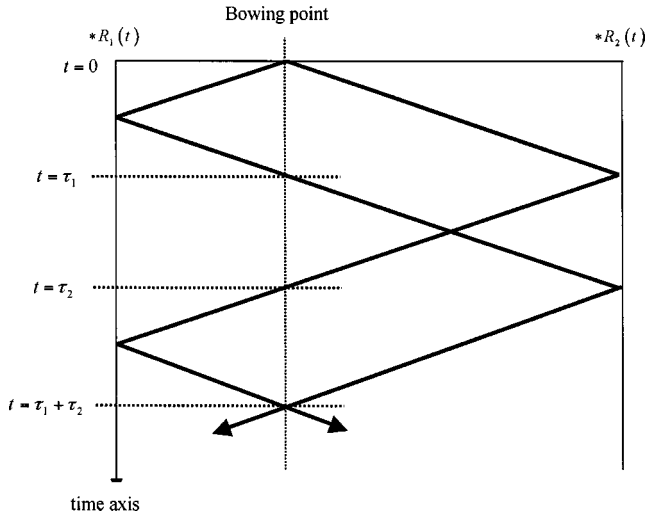


FIG. 1. Propagation–reflection process in a free uniform string.

where $f_0(t)=h_0(t)*g(t)$.

This equation is valid for any time t. If we write $f(t-\tau)$ and combine it with Eq. (4), we obtain the general convolution for wind instruments

$$f(t)-f_0(t)=R_2(t)*\sigma_0(t)*[R_1(t)*f(t-\tau)+f_0(t-\tau)]. \quad (5)$$

B. String instruments description

For the string family, formulations close to the one presented in the previous section can be found in the literature (see, for example, Friedlander, 1953; McIntyre and Woodhouse, 1979; McIntyre *et al.*, 1983).

The propagation–reflection rationale of the initial upward and downward waves leading to $h(t)$ is depicted in Fig. 1. The general convolution equation in this case is

$$f(t)-\tfrac{1}{2}f_0(t)=R_1(t)*R_2(t)*\sigma_1(t)*\sigma_2(t)*[f(t-\tau)+\tfrac{1}{2}f_0(t-\tau)]+\tfrac{1}{2}[R_1(t)*\sigma_1(t)*f_0(t-\tau_1)+R_2(t)*\sigma_2(t)*f_0(t-\tau_2)], \quad (6)$$

where $R_1(t)$ and $R_2(t)$ are the elementary reflection functions associated with the two ends, τ_1 and τ_2 the two characteristic times and $\tau=\tau_1+\tau_2$.

III. GENERAL CONSIDERATIONS ON DISSIPATION AND DISPERSION MECHANISMS

Some considerations have to be made concerning the dissipation and dispersion mechanisms before formulating them more precisely.

For wind instruments, the energy losses are associated with internal frictions (air–wall friction, air viscosity, and thermal exchanges) and radiation at the open ends (the latter corresponding to the general category of ''energy transfer to other systems''). For the string family, they also come from internal damping and energy transfer through the supports. In any case, internal frictions have been represented until now by the damping function $\sigma_i(t)$, and energy transfer to other systems, by the local reflection functions $R_i(t)$.

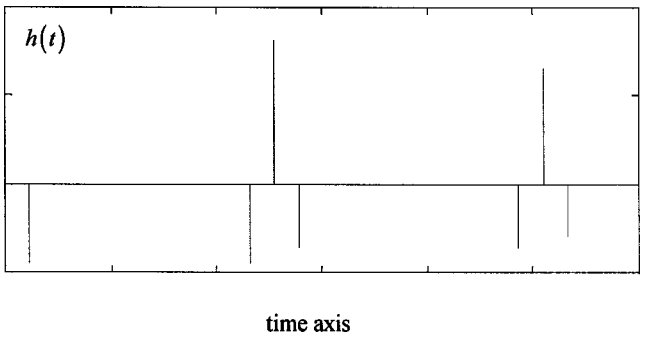


FIG. 2. Qualitative impulse response for a string with nondispersive losses localized at the ends.

The exact formulation of those phenomena is complicated and can be found, for example, in Polack *et al.* (1987), Valette and Cuesta (1993), and Fletcher and Rossing (1991). In general, all those mechanisms imply an intensity decrease whose rate is frequency dependent (and so leads to dispersion effects).

An intensity decrease can be represented in the time domain in many different ways. The simplest one consists in a multiplication by a factor R with modulus lower than unity [or, what is the same, in a convolution with a Dirac distribution with intensity lower than unity, $R(t)=R\,\delta(t)$, $|R|<1$]. This oversimplified formulation was used by Maganza *et al.* to study the possible behaviors of a clarinet-like system (Maganza *et al.*, 1986). For the case of the string, this kind of approach leads to the impulse response shown in Fig. 2, as was first studied by Raman (1918).

A more realistic model should take into account the effect of dispersion, that is, the frequency-dependent wave speed. From a qualitative point of view, dispersion is responsible not only for an amplitude decrease of the physical magnitudes but also for a time spreading. There are different physical reasons for this effect. In general, losses generate dispersion. In a 1D system, the nonuniformity of the material characteristics (density, elasticity...) implies dispersion. In strings, the bending stiffness is the main reason for the time spreading of the signals (Woodhouse, 1993).

Accurate mathematical models for dispersion associated with these various reasons have been obtained in the past by many authors. However, in order to simplify the mathematical manipulations and make possible the simple differential formulations we seek, we have chosen to describe them systematically with just an exponential kernel whose amplitude and exponent can be adapted to better fit the different cases. The convolution with such kernels spreads and attenuates the signals.

Thus, a dispersive energy transfer (radiation from open ends in the case of wind instruments, transfer through the supports for strings) will be represented by local reflection functions $R(t)\propto\exp(-\alpha t)$ (with α constant). Dispersive exponential reflection functions for the case of woodwind-like systems were already presented in Barjau *et al.* (1997a). If internal losses are neglected, this kind of formulation for the case of a string leads to the impulse response depicted in Fig. 3.

When dispersion is not localized (as in the case of en-

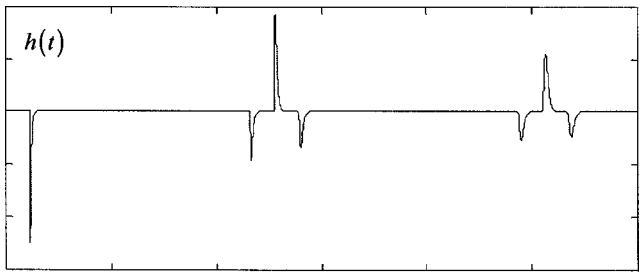


FIG. 3. Qualitative impulse response for a string with dispersive losses localized at the ends.

ergy transfer just described) but takes place during propagation, it leads to qualitatively different impulse responses as far as the early response $h_0(t)$ is concerned. For the string family, different shapes of $h_0(t)$ are discussed in McIntyre and Woodhouse (1979).

In order to retain a certain simplicity in the mathematical formulation, the damping function for the extended dispersive internal losses will be treated in a simplified way: for each basic span length L_i (L_0 for the case of woodwinds and L_1, L_2 for the case of strings), the damping function will be taken to be $\sigma_i(t)=\beta_i \exp(-\tilde{\beta}_i t)$, with β_i, $\tilde{\beta}_i \propto 1/L_i$ [but $\beta_i < \tilde{\beta}_i$ in order to have $\int \sigma_i(t)dt<1$]. For longer trips NL_i, the global effect of damping will be obtained through N convolutions $(*\sigma_i(t))^N$.

The impulse response with internal losses that will be used for the string case is shown in Fig. 4.

IV. WIND INSTRUMENTS WITH NONDISPERSIVE LOSSES AT THE ENDS

Though we have worked with general input and output variables, $g(t)$ and $f(t)$, respectively, and general wind and string instruments up till now, we will restrict ourselves to the case of woodwinds from now on [and so $g(t)=\nu(t)$ will be the air velocity, and $f(t)=p(t)$ will be the acoustical pressure, both at the bore input section, or vice versa]. The equivalent results for bowed strings [where $g(t)=F(t)$ is the friction force between bow and string, and $f(t)=\nu(t)$ is the transverse string velocity at the bowing point] will be summarized in Sec. IV and Table II.

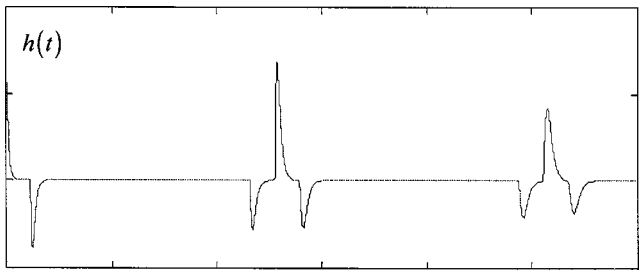


FIG. 4. Qualitative impulse response for a string including non localized dispersive effects and nondispersive losses at the ends.

A. Cylindrical wind instruments

The particular form of the convolution in Eq. (5) can be obtained either by taking $f(t)=p(t)$ or $f(t)=\nu(t)$. If the acoustical pressure is taken as the output variable, the impulse response $h(t)$ leading to Eq. (5) is the pressure response to an impulsive velocity $\nu(t)$. A physical realization of a velocity impulse is that of suddenly tapping and closing the input section. The fact of not reopening it guarantees a zero value for $\nu(t)$ once the impulse has been generated.

That means that the pressure reflection function at the input section, $R_1(t)$, must correspond to a closed rigid-end condition, and so $R_1(t)=R_{1p}^{\mathrm{ce}}(t)=\delta(t)$ (where the superscript ce stands for ''closed end'').

If the air velocity $\nu(t)$ is chosen as the output variable, the impulse response is the result of an impulsive pressure at $t=0$. In this case, as the simplest way to guarantee a zero pressure value for $t>0$ is to keep the input section open, the velocity reflection function $R_1(t)$ must correspond to an idealized open-end condition, and so $R_1(t)=R_{1\nu}^{\mathrm{oe}}(t)=\delta(t)$ (where the superscript oe stands for ''open end'').

If we take $g(t)=\nu(t)$ and $f(t)=p(t)$, the particularization of Eq. (5) to a cylindrical bore with an open output end gives

$$p(t)-p_0(t)=R_{2p}^{\mathrm{oe}}(t)*[R_{1p}^{\mathrm{ce}}(t)*p(t-\tau)+p_0(t-\tau)], \tag{7}$$

with $p_0(t)=Z_0\nu(t)$, $R_{2p}^{\mathrm{oe}}(t)=-R_2\delta(t)$ (with $R_2>0$) and $R_{1p}^{\mathrm{ce}}(t)=\delta(t)$. Finally

$$p(t)-Z_0\nu(t)=-R_2[p(t-\tau)+Z_0\nu(t-\tau)]. \tag{8}$$

A straightforward calculation shows that Eq. (8) still holds if we take $g(t)=p(t)$ and $f(t)=\nu(t)$. Equation (8) is identical to the one obtained by Maganza *et al.* (1986) for their simplified model of a clarinet.

The case of a cylindrical bore with a closed output end can be obtained in the same way. The equivalent to Eq. (8) is

$$p(t)-p_0(t)=R_{2p}^{\mathrm{ce}}(t)*[R_{1p}^{\mathrm{ce}}(t)*p(t-\tau)+p_0(t-\tau)], \tag{9}$$

where $p_0(t)=Z_0\nu(t)$, $R_{2p}^{\mathrm{ce}}(t)=R_2\delta(t)$ (with $0<R_2<1$ in order to allow a certain dissipation associated to this end) and $R_{1p}^{\mathrm{ce}}(t)=\delta(t)$.

From a strict point of view, Eqs. (8) and (9) cannot present any periodic behavior if $R_2<1$. However, the extreme case of perfect reflecting (or lossless) ends (where $R_2=1$) allows some qualitative comments on possible periodic behaviors.

For the case of an idealized open output end, Eq. (8) becomes

$$p(t)-Z_0\nu(t)=-p(t-\tau)-Z_0\nu(t-\tau), \tag{10}$$

while Eq. (9), describing the cylindrical bore with closed output end, becomes

$$p(t)-Z_0\nu(t)=p(t-\tau)+Z_0\nu(t-\tau). \tag{11}$$

The pressure-dependent terms suggest a periodicity 2τ for the open output end system and a periodicity τ for the closed output end, while the velocity-dependent ones suggest exactly the contrary. The actual behavior will then be given

by the relative importance of those two kinds of terms.

Let's assume that we are dealing with a reed woodwind. It is clear that the pressure terms in Eqs. (10) and (11) are related to the coupling mechanism (Barjau and Agulló, 1989; Fletcher and Rossing, 1991) or, in other words, the action of the air column on the reed. The velocity terms are related to the reed dynamics. The more intense the Bernoulli mechanism, the more important the velocity terms.

From those considerations, some general conclusions can be extracted. A cylindrical reed instrument with a low Bernoulli mechanism [and so with moderate values for $\nu(t)$] would tend to show a 2τ-periodic behavior if its output end was open (and a τ-periodic behavior if it was closed). However, if the Bernoulli effect is high (as is the case, for example, in double reeds) or, in general, if the nonlinearity gives high values for $\nu(t)$, the open cylinder could show a τ-periodicity (and the closed one could have a 2τ-periodicity).

B. Conical wind instruments

The case of a conical bore is more involved. On one hand, the initial pressure signal and the input flow are not simply proportional (Martinez and Agulló, 1988), but

$$p_0(t)=Z_0\left[\delta(t)-H(t)\frac{c}{r_0}\exp(-ct/r_0)\right]*\nu(t), \tag{12}$$

where $H(t)$ is the Heaviside step distribution, c is the sound speed in air, and r_0 is the distance between the input section and the full cone apex ($r_0>0$ for divergent cones—that is, with an output radius larger than the input one, and $r_0<0$ for convergent ones).

On the other hand, even if the pressure reflection function for open ends is the same as in the cylindrical case, $R_p^{\mathrm{oe}}(t)=-R\,\delta(t)$ with $0<R<1$, for closed rigid ends it is totally different (Martinez and Agulló, 1988)

$$R_p^{\mathrm{ce}}(t)=\delta(t)-2\frac{c}{r_0}H(t)\exp(-ct/r_0). \tag{13}$$

As seen in the previous subsection, it is possible to work with $f(t)=p(t)$ and $g(t)=\nu(t)$, or with $f(t)=\nu(t)$ and $g(t)=p(t)$. In the present case, we will restrict the calculations to the first choice. Thus, for a conical bore with an output open end, Eq. (5) becomes

$$\begin{aligned}p(t)-Z_0&\left[\delta(t)-\frac{c}{r_0}H(t)\exp(-ct/r_0)\right]*\nu(t)\\&=R_{2p}^{\mathrm{oe}}(t)*\left[R_{1p}^{\mathrm{ce}}(t)*p(t-\tau)\right.\\&\quad\left.+Z_0\left\{\delta(t)-\frac{c}{r_0}H(t)\exp(-ct/r_0)\right\}*\nu(t-\tau)\right],\end{aligned} \tag{14}$$

with $R_{2p}^{\mathrm{oe}}(t)=-R_2\,\delta(t)$ $(0<R_2<1)$, and $R_{1p}^{\mathrm{ce}}(t)=\delta(t)-2(c/r_0)H(t)\exp(-ct/r_0)$.

Substitution of the reflection functions in Eq. (14) leads to

$$\begin{aligned}p(t)-Z_0\nu(t)+Z_0&\frac{c}{r_0}H(t)\exp(-ct/r_0)*\nu(t)\\&=-R_2\left[p(t-\tau)+Z_0\nu(t-\tau)-\frac{c}{r_0}H(t)\exp(-ct/r_0)\right.\\&\quad\left.*\{2p(t-\tau)+Z_0\nu(t-\tau)\}\right].\end{aligned} \tag{15}$$

This formulation is no longer a simple iterative one as in the case of the cylinder. The use of a direct and an inverse Laplace transform in Eq. (15) leads to the differential formulation of the problem

$$\begin{aligned}\frac{d}{dt}&[p(t)-Z_0\nu(t)]+\frac{c}{r_0}p(t)\\&=-R_2\left\{\frac{d}{dt}[p(t-\tau)+Z_0\nu(t-\tau)]-\frac{c}{r_0}p(t-\tau)\right\}.\end{aligned} \tag{16}$$

It is worth pointing out that the corresponding cylindrical case [Eq. (8)] can be obtained from Eq. (16) by taking $r_0\to\infty$.

There is an alternative formulation of the conical problem. If we define

$$\begin{aligned}p_1(t)&\equiv Z_0\frac{c}{r_0}H(t)\exp(-ct/r_0)*\nu(t)\\p_2(t-\tau)&\equiv\frac{c}{r_0}H(t)\exp(-ct/r_0)*p(t-\tau),\end{aligned} \tag{17}$$

Eq. (15) can be rewritten as

$$\begin{aligned}p(t)-Z_0\nu(t)+p_1(t)=&-R_2[p(t-\tau)+Z_0\nu(t-\tau)\\&-2p_2(t-\tau)-p_1(t-\tau)].\end{aligned} \tag{18}$$

This algebraic equation has to be completed with two first-order differential equations governing the new variables

$$\begin{aligned}\frac{dp_1(t)}{dt}&=-\frac{c}{r_0}[p_1(t)-Z_0\nu(t)],\\\frac{dp_2(t)}{dt}&=-\frac{c}{r_0}[p_2(t)-p(t)].\end{aligned} \tag{19}$$

The mathematical manipulation leading to the delayed differential equation for the closed output end case is much longer and tedious. As there are no conical woodwinds with closed output end, this case will not be developed. From now on, only open output end woodwinds will be considered.

The lossless cases ($R_2=1$) may allow some considerations of possible periodic solutions. After reorganization, Eq. (16) becomes

$$\begin{aligned}\frac{dp(t)}{dt}+\frac{c}{r_0}p(t)-Z_0\frac{d\nu(t)}{dt}=&-\frac{dp(t-\tau)}{dt}+\frac{c}{r_0}p(t-\tau)\\&-Z_0\frac{d\nu(t-\tau)}{dt}.\end{aligned} \tag{20}$$

This equation is far more complicated than those describing the cylindrical cases. In order to analyze the possible implicit periodicities, there are now three different kinds of

TABLE I. Woodwinds.

Change of variables	Compact formulation
Nondispersive losses at the ends	
Cylinder: $x(t)=p(t)$, $y(t)=Z_0\nu(t)$ Cone: $x(t)=[dp(t)]/dt$, $y(t)=Z_0[d\nu(t)]/dt-(c/r_0)p(t)$	$x(t)-y(t)=-R_2\{x(t-\tau)+y(t-\tau)\}$
Dispersive losses at the ends	
Cylinder: $x(t)=p(t)$, $y(t)=Z_0\nu(t)$ Cone: $x(t)=[dp(t)]/dt$, $y(t)=Z_0[d\nu(t)]/dt-(c/r_0)p(t)$	$[(d/dt)+\tilde{\alpha}]\{x(t)-y(t)\}=-\alpha\{x(t-\tau)+y(t-\tau)\}$
Dispersive internal losses	
Cylinder: $x(t)=[(d/dt)+\tilde{\eta}]p(t)$, $y(t)=Z_0\eta\nu(t)$ Cone: $x(t)=[(d/dt)+\tilde{\eta}][dp(t)]/dt$, $y(t)=Z_0\eta[d\nu(t)]/dt-(c/r_0)[(d/dt)+\tilde{\eta}]p(t)$	$[(d/dt)+\tilde{\beta}_0]\{x(t)-y(t)\}=-R_2\beta_0\{x(t-\tau)+y(t-\tau)\}$
Dispersive internal losses and dispersive losses at the ends	
Cylinder: $x(t)=[(d/dt)+\tilde{\eta}]p(t)$, $y(t)=Z_0\eta\nu(t)$ Cone: $x(t)=[(d/dt)+\tilde{\eta}][dp(t)]/dt$, $y(t)=Z_0\eta[d\nu(t)]/dt-(c/r_0)[(d/dt)+\tilde{\eta}]p(t)$	$[(d/dt)+\tilde{\alpha}][(d/dt)+\tilde{\beta}_0]\{x(t)-y(t)\}=-\alpha\{x(t-\tau)+y(t-\tau)\}$

terms whose relative importance has to be assessed, the third and new one being that of the terms containing the cone truncation (here represented through r_0).

A first interesting case is that of a full cone ($r_0=0$). The dominating terms in Eq. (20) are the nonderivated pressures

$$p(t)=p(t-\tau). \tag{21}$$

This implies that, independently from the strength of the nonlinearity, the conical air column will impose a τ periodicity to the output variable $p(t)$ [and so to $\nu(t)$ as well]. This is nothing more than the classical result obtained in the frequency domain through modal analysis and more difficult to justify in the time domain.

For truncated cones, where $r_0\neq 0$, the resulting periodicity depends on the strength of the nonlinearity $\nu(t)=\mathrm{NL}[p(t)]$ (for the case of a double-reed conical woodwind, that means the strength of the Bernoulli effect). If the nonlinear terms are the dominant ones, Eq. (20) indicates again a τ-periodicity.

If the velocity terms are negligible, Eq. (20) becomes

$$\left[\frac{d}{dt}+\frac{c}{r_0}\right]p(t)=-\left[\frac{d}{dt}-\frac{c}{r_0}\right]p(t-\tau). \tag{22}$$

Again, an assessment of the degree of truncation is needed. For high truncations, one can assume that the terms multiplied by (c/r_0) are small compared to the others. In this case, Eq. (22) implies a 2τ periodicity. For intermediate truncations, the term derived in Eq. (20) suggests a 2τ periodicity, while the ''truncation'' term suggests a τ periodicity. So, nothing definite can be concluded [actually, some conical instruments belonging to the bagpipe family produce sounds whose frequency lies between $1/2\tau$ and $1/\tau$, something which points exactly in this direction; see Jeltsch and Gibiat (2000)].

V. WIND INSTRUMENTS WITH DISPERSIVE LOSSES AT THE ENDS

The delayed equations for cylindrical and conical woodwinds with open output section and dispersive losses at the ends can be obtained from Eqs. (8) and (16), respectively, by substituting the simple scalar multiplication by $-R_2$ by a convolution product with the dispersive open-end reflection function $R_2(t)=-\alpha\exp(-\tilde{\alpha}t)$. Again, the convolution can be eliminated by performing a direct and an inverse Laplace transform. The corresponding differential equations for the output end cylinder and the output end cone are given in Table I. They coincide with those appearing in Barjau *et al.* (1997a), as expected.

VI. WIND INSTRUMENTS WITH INTERNAL DISPERSIVE LOSSES

As discussed in Sec. III, if internal losses are taken into account the $f(t)$ early response to an impulsive condition in $g(t)$, $f_0(t)$, will contain an extra wake that can be represented mathematically through a new convolution with an exponential kernel. If we call $f_0^{il}(t)$ the early $f(t)$ response with internal losses and, $f_0(t)$ that corresponding to the lossless case, then

$$f_0^{il}(t)=[\delta(t)+\theta(t)]*f_0(t), \tag{23}$$

with $\theta(t)=H(t)\eta\exp(-\tilde{\eta}t)$ (with $\tilde{\eta}>0$). The dissipation kernel $\sigma_0(t)$ appearing at the right-hand side of Eq. (5) corresponds to a round trip along the bore and it is not in principle equal to $\theta(t)$. It will be written as $\sigma_0(t)$

TABLE II. String instruments.

Change of variables	Compact formulation
Nondispersive losses at the ends	
$x(t)=\nu(t),\ y(t)=\frac{1}{2}A_0F(t)$	$x(t)-y(t)=R_1R_2[x(t-\tau)+y(t-\tau)]-[R_1y(t-\tau_1)+R_2y(t-\tau_2)]$
Dispersive losses at the ends	
$x(t)=\nu(t),\ y(t)=\frac{1}{2}A_0F(t)$	$[(d/dt)+\tilde{\alpha}_1][(d/dt)+\tilde{\alpha}_2]\{x(t)-y(t)\}$ $=\alpha_1\alpha_2[x(t-\tau)+y(t-\tau)]-[\alpha_1[(d/dt)+\tilde{\alpha}_2]y(t-\tau_1)$ $+\alpha_2[(d/dt)+\tilde{\alpha}_1]y(t-\tau_2)]$
Dispersive internal losses	
$x(t)=[(d/dt)+\tilde{\eta}]\nu(t),\ y(t)=\frac{1}{2}A_0\,\eta F(t)$	$[(d/dt)+\tilde{\beta}_1][(d/dt)+\tilde{\beta}_2]\{x(t)-y(t)\}$ $=R_1R_2[x(t-\tau)+y(t-\tau)]-[R_1\beta_1[(d/dt)+\tilde{\beta}_2]y(t-\tau_1)$ $+R_2\beta_2[(d/dt)+\tilde{\beta}_1]y(t-\tau_2)]$
Dispersive internal losses and dispersive losses at the ends	
$x(t)=[(d/dt)+\tilde{\eta}]\nu(t),\ y(t)=\frac{1}{2}A_0\,\eta F(t)$	$[(d/dt)+\tilde{\alpha}_1][(d/dt)+\tilde{\alpha}_2][(d/dt)+\tilde{\beta}_1][(d/dt)+\tilde{\beta}_2]\{x(t)-y(t)\}$ $=\alpha_1\alpha_2[x(t-\tau)+y(t-\tau)]-[\alpha_1\beta_1[(d/dt)+\tilde{\alpha}_2]$ $\times[(d/dt)+\tilde{\beta}_2]y(t-\tau_1)+\alpha_2\beta_2[(d/dt)+\tilde{\alpha}_1]$ $\times[(d/dt)+\tilde{\beta}_1]y(t-\tau_2)]$

$=H(t)\beta_0\exp(-\tilde{\beta}_0 t)$. The delayed differential models for both nondispersive and dispersive reflection functions $R_i(t)$ are shown in Table I.

VII. MODELS FOR STRING INSTRUMENTS

The same kind of models that have been developed for wind instruments can be obtained for the string family. Though the simplest one (with nondispersive losses concentrated at the end supports) is already known (see, for example, Raman, 1918), the differential formulations including dispersive effects cannot be found in scientific literature.

The same kind of notation will be used: $R_{1,2}(t)$ will be the reflection functions at the ends ($R_{1,2}(t)=-R_{1,2}\delta(t)$, with $0<R_{1,2}<1$ for the nondispersive case, and $R_{1,2}(t)=-\alpha_{1,2}\exp(-\tilde{\alpha}_{1,2}t)$ for the dispersive one); $\theta(t)$ will represent the initial spreading of the signal [$\theta(t)=H(t)\eta\times\exp(-\tilde{\eta}t)$], and $\sigma_{1,2}(t)$ will be the kernel containing the internal losses [$\sigma_{1,2}(t)=H(t)\beta_{1,2}\exp(-\tilde{\beta}_{1,2}t)$, where the subscripts refer to the two string spans].

The detailed calculations will be skipped, but the corresponding differential equations are shown in Table II.

VIII. GENERAL DISCUSSION ON WIND INSTRUMENT MODELS

The two families of musical instruments that have been studied have an important common characteristic but a main difference. On one hand, they have all been modeled as 1D systems; on the other hand, the string family has the excitation point located somewhere between the two system ends while the woodwinds have the excitation device located at one end. This difference is the reason for having more complicated equations for strings than for woodwinds.

For each family, we have developed different models leading to apparently different behaviors (as the final equations are not the same). However, it is easy to see that there is always a variable transformation that allows the equations to be written under the same general structure (within the same family).

For cylindrical and conical woodwinds, Table I presents such transformations and the resulting compact form of the equations for all the models. The same thing is presented in Table II for strings.

For the case of woodwinds, a further step in compaction is still possible. The two different formulations appearing in Table I can be expressed in a very abstract form as

$$D[x(t)-y(t)]=x(t-\tau)+y(t-\tau), \tag{24}$$

where D is a general differential operator. This equation allows a very simple interpretation. The time delay appearing at the right-hand side is of course related to the propagation time. Thus, the right-hand side represents in a very general way a downward wave [$\psi_{\text{down}}(t)$] reaching the input section, while the left-hand side corresponds to an upward wave [$\psi_{\text{up}}(t)$]. The differential operator represents then the changes undergone by the initial downward wave due to dissipation and reflection processes: $D[\psi_{\text{up}}(t)]=\psi_{\text{down}}(t-\tau)$ (of course, it could be placed at the right-hand side as an integrator, $\psi_{\text{up}}(t)=I[\psi_{\text{down}}(t-\tau)]$, but to be coherent with the philosophy of the paper, the differential formulation will be kept). The fact that the differential operator appears just at the left-hand side implies that both waves are totally un-

coupled. In other words, they correspond necessarily to the eigenvariables of the problem.

Before presenting in detail the eigenfunctions of the conservative and dissipative cases, the general formulation in Eq. (24) still deserves some comments. The variable $x(t)$ is essentially a linear function of the pressure and its two first-time derivatives: $x(t)=L[p(t),\dot{p}(t),\ddot{p}(t)]$, while $y(t)$ depends on the same variables but is essentially nonlinear: $y(t)=NL[p(t),\dot{p}(t),\ddot{p}(t)]$. Thus, in general both the upward and the downward waves, $\psi_{\text{up}}(t)\equiv x(t)-y(t)$ and $\psi_{\text{down}}(t)\equiv x(t)+y(t)$, respectively, are nonlinear functions. However, it may be possible to find a different nonlinear dependence between $\psi_{\text{down}}(t)$ and $\psi_{\text{up}}(t)$, $\psi_{\text{down}}(t)=NL[\psi_{\text{up}}(t)]$. In that case, Eq. (24) belongs to a well-known class of delayed equations describing many different phenomena in various fields (see, for example, Manneville, 1990; Murray, 1993)

$$D[\psi_{\text{up}}(t)]=\text{NL}[\psi_{\text{up}}(t-\tau)]. \tag{25}$$

A systematic study of fixed points, stability, and possible limit cycles can be done once the basic nonlinearity has been formulated.

For the case of a nondissipative cylindrical bore, it is well known that the upward and downward waves able to propagate without interfering can be expressed from the usual acoustical variables as

$$\psi_{\text{up}}^{\text{cyl}}(t)=p(t)-Z_0\nu(t), \tag{26}$$

$$\psi_{\text{down}}^{\text{cyl}}(t)=p(t)+Z_0\nu(t). \tag{27}$$

For nondissipative conical air columns, the expression of the independent upward and downward waves is far more complicated (Meynial, 1987; Agulló *et al.*, 1988)

$$\psi_{\text{up}}^{\text{cone}}(t)=p(t)-Z_0\nu(t)+\frac{c}{r_0}\int_0^t p(t')dt', \tag{28}$$

$$\psi_{\text{down}}^{\text{cone}}(t)=p(t)+Z_0\nu(t)-\frac{c}{r_0}\int_0^t p(t')dt'. \tag{29}$$

These independent waves can also be expressed in a differential way by just deriving the previous equations

$$\frac{d}{dt}\psi_{\text{up}}^{\text{cone}}(t)=\frac{dp(t)}{dt}-Z_0\frac{d\nu(t)}{dt}+\frac{c}{r_0}p(t), \tag{30}$$

$$\frac{d}{dt}\psi_{\text{down}}^{\text{cone}}(t)=\frac{dp(t)}{dt}+Z_0\frac{d\nu(t)}{dt}-\frac{c}{r_0}p(t). \tag{31}$$

If Eqs. (26) and (27) are used, these waves can be rewritten as

$$\frac{d}{dt}\psi_{\text{up}}^{\text{cone}}(t)=\left[\frac{d}{dt}+\frac{c}{2r_0}\right]\psi_{\text{up}}^{\text{cyl}}(t)+\frac{c}{2r_0}\psi_{\text{down}}^{\text{cyl}}(t), \tag{32}$$

$$\frac{d}{dt}\psi_{\text{down}}^{\text{cone}}(t)=\left[\frac{d}{dt}-\frac{c}{2r_0}\right]\psi_{\text{down}}^{\text{cyl}}(t)-\frac{c}{2r_0}\psi_{\text{up}}^{\text{cyl}}(t), \tag{33}$$

and so

$$\psi_{\text{up}}^{\text{cone}}(t)=\psi_{\text{up}}^{\text{cyl}}(t)+\frac{c}{2r_0}\int_0^t[\psi_{\text{up}}^{\text{cyl}}(t')+\psi_{\text{down}}^{\text{cyl}}(t')]dt', \tag{34}$$

$$\psi_{\text{down}}^{\text{cone}}(t)=\psi_{\text{down}}^{\text{cyl}}(t)-\frac{c}{2r_0}\int_0^t[\psi_{\text{down}}^{\text{cyl}}(t')+\psi_{\text{up}}^{\text{cyl}}(t')]dt'. \tag{35}$$

These new expressions allow an interesting comparison between the behavior of cylindrical and conical bores. The asymmetry in the sign of the terms containing c/r_0 is simply due to the asymmetrical aspect of a cone when going downwards or upwards: in one case it is divergent/convergent, while it is convergent/divergent in the other.

Two other aspects merit comment. On one hand, the integration appearing at the right-hand side of both equations means that a conical entrance spreads the signals that would be able to propagate without changing their shape in a cylindrical bore. On the other hand, the fact that both equations include the two propagative cylindrical waves means that a cone can be seen as a tube where a plane wave would undergo a continuous reflection–transmission process. It can thus be assimilated to a set of assembled cylindrical tubes of different diameters (as already done by many authors, see, for example, Bruneau, 1983).

The existence of dispersive internal dissipations in air columns is responsible for a change in shape of these independent waves. Thus, the eigenvariables cannot be the same. Even if their calculation is not evident, the general form in Eq. (24) allows a straightforward identification.

For the cylindrical bore, internal dispersive losses lead to

$$\frac{d}{dt}\tilde{\psi}_{\text{up}}^{\text{cyl}}(t)=\frac{1}{2}\frac{d}{dt}[\psi_{\text{up}}^{\text{cyl}}(t)+\psi_{\text{down}}^{\text{cyl}}(t)]+\tfrac{1}{2}[(\tilde{\eta}+\eta)\psi_{\text{up}}^{\text{cyl}}(t)+(\tilde{\eta}-\eta)\psi_{\text{down}}^{\text{cyl}}(t)], \tag{36}$$

$$\frac{d}{dt}\tilde{\psi}_{\text{down}}^{\text{cyl}}(t)=\frac{1}{2}\frac{d}{dt}[\psi_{\text{down}}^{\text{cyl}}(t)+\psi_{\text{up}}^{\text{cyl}}(t)]+\tfrac{1}{2}[(\tilde{\eta}+\eta)\psi_{\text{down}}^{\text{cyl}}(t)+(\tilde{\eta}-\eta)\psi_{\text{up}}^{\text{cyl}}(t)]. \tag{37}$$

These equations share two aspects with the nondissipative conical ones: the signals are also spread from the very beginning, and some kind of continuous reflection–transmission process goes on. However, a cylinder with internal losses will never be totally equivalent to a conservative cone because the basic asymmetry of divergence/convergence, essential in the latter case, will never be present in the former one. If an analogous conservative system had to be found, a cylindrical bore with equally spaced diaphragms would be a possibility, as it would contain elements responsible for the reflection–transmission phenomena and would be symmetrical for downwards or upwards propagation.

The propagative eigenfunctions for the dissipative cone, expressed as a function of nondissipative waves, are

$$\frac{d}{dt}\tilde{\psi}_{\mathrm{up}}^{\mathrm{cone}}(t)=\frac{1}{2}\left[\frac{d}{dt}+\tilde{\eta}\right]\left[\frac{d}{dt}+\frac{c}{r_0}\right]\left[\psi_{\mathrm{up}}^{\mathrm{cone}}(t)+\psi_{\mathrm{down}}^{\mathrm{cone}}(t)\right]+\frac{\eta}{2}\frac{d}{dt}\{\psi_{\mathrm{up}}^{\mathrm{cyl}}(t)-\psi_{\mathrm{down}}^{\mathrm{cyl}}(t)\},\tag{38}$$

$$\frac{d}{dt}\tilde{\psi}_{\mathrm{down}}^{\mathrm{cone}}(t)=\frac{1}{2}\left[\frac{d}{dt}+\tilde{\eta}\right]\left[\frac{d}{dt}-\frac{c}{r_0}\right]\left[\psi_{\mathrm{down}}^{\mathrm{cone}}(t)+\psi_{\mathrm{up}}^{\mathrm{cone}}(t)\right]+\frac{\eta}{2}\frac{d}{dt}\{\psi_{\mathrm{down}}^{\mathrm{cyl}}(t)-\psi_{\mathrm{up}}^{\mathrm{cyl}}(t)\}.\tag{39}$$

The same comments that have made done about Eqs. (34) and (35) can be made now. The internal losses are responsible for a spread of the conservative waves plus a continuous reflection–transmission process.

If we rewrite $\tilde{\psi}_{\mathrm{up}}^{\mathrm{cone}}(t)$ and $\tilde{\psi}_{\mathrm{down}}^{\mathrm{cone}}(t)$ in terms of the dissipative cylindrical waves

$$\frac{d}{dt}\tilde{\psi}_{\mathrm{up}}^{\mathrm{cone}}(t)=\frac{d^2}{dt^2}\tilde{\psi}_{\mathrm{up}}^{\mathrm{cyl}}(t)+\frac{c}{2r_0}\frac{d}{dt}\left[\tilde{\psi}_{\mathrm{up}}^{\mathrm{cyl}}(t)+\tilde{\psi}_{\mathrm{down}}^{\mathrm{cyl}}(t)\right],\tag{40}$$

$$\frac{d}{dt}\tilde{\psi}_{\mathrm{down}}^{\mathrm{cone}}(t)=\frac{d^2}{dt^2}\tilde{\psi}_{\mathrm{down}}^{\mathrm{cyl}}(t)-\frac{c}{2r_0}\frac{d}{dt}\left[\tilde{\psi}_{\mathrm{down}}^{\mathrm{cyl}}(t)+\tilde{\psi}_{\mathrm{up}}^{\mathrm{cyl}}(t)\right],\tag{41}$$

the analogy between a cone and a set of cylindrical spans with different diameters becomes evident again.

IX. CONCLUSIONS

We have been able to establish a general differential formulation for simplified woodwinds (cylindrical and conical) that includes different models for losses (localized and spread—internal-, dispersive, and nondispersive). We have shown that this formulation has a very intuitive interpretation, and from it we have been able to do a systematic determination of the uncoupled variables (or eigenvariables) for each model. It has also been possible to establish a comparison between cylinders and cones, and to find alternative models sharing the same qualitative behavior.

The case of strings has also been considered, and a general differential formulation has also been found that includes four different models for losses. However, the existence of three different time delays in the equations does not allow the determination of the uncoupled variables.

Agulló, J., and Barjau, A. (**1986**). "The reflection function: A matrix approach versus FFT," J. Sound Vib. **106**(2), 193–201.

Agulló, J., Barjau, A., and Martínez, J. (**1988**). "Alternatives to the impulse response $h(t)$ to describe the acoustical behavior of conical ducts," J. Acoust. Soc. Am. **84**(5), 1606–1612.

Antunes, J., Tafasca, M., and Borsoi, L. (**2000**). "Simulation des Régimes Vibratoires Non-linéaires d'une Corde de Violon," Rev. Fr. Mécan. **3**, 193–202.

Barjau, A., and Agulló, J. (**1989**). "Calculation of the starting transients of a double reed conical woodwind," Acustica **60**(5), 204–210.

Barjau, A., Gibiat, V., and Grand, N. (**1997a**). "Study of woodwind-like systems through nonlinear differential equations. I. Simple geometry," J. Acoust. Soc. Am. **102**(5), 3023–3031.

Barjau, A., and Gibiat, V. (**1997b**). "Study of woodwind-like systems through nonlinear differential equations. II. Real geometry," J. Acoust. Soc. Am. **102**(5), 3032–3037.

Bruneau, M. (**1983**). *Introduction aux Théories de L'acoustique* ("Introduction to Theories on Acoustics") (Publications de l'Université du Maine, Le Mans), pp. 78–82.

Fletcher, N. H., and Rossing, T. D. (**1991**). *The Physics of Musical Instruments* (Springer, New York).

Friedlander, F. G. (**1953**). "On the oscillations of the bowed string," Proc. Cambridge Philos. Soc. **49**, 516–530.

Gibiat, V. (**1988**). "Phase space representations of acoustical musical signals," J. Sound Vib. **123**, 529–536.

Jeltsch, J., and Gibiat, V. (**2000**). "Le fonctionnement acoustique des cornemuses à travers les exemples de la Boha Landaise et de la chabrette limousine," Actes du 5ième Congres Français d'Acoustique, CFA 2000, edited by Presses Polytechniques et Universitaires Romandes.

Maganza, C., Caussé, R., and Laloë, F. (**1986**). "Bifurcations, period doublings, and chaos in clarinet-like systems," Europhys. Lett. **1**, 295–299.

Manneville, P. (**1990**). *Dissipative Structures and Weak Turbulence* (Academic, New York).

Martínez, J., and Agulló, J. (**1988**). "Conical bores. I. Reflection functions associated with discontinuities," J. Acoust. Soc. Am. **84**, 1613–1619.

McIntyre, M., Schumacher, R., and Woodhouse, J. (**1983**). "On the oscillations of musical instruments," J. Acoust. Soc. Am. **74**, 1325–1345.

McIntyre, M., and Woodhouse, J. (**1979**). "On the fundamentals of bowed-string dynamics," Acustica **43**(2), 93–108.

Meynial, X. (**1987**). "Systèmes micro-intervalles pour instruments à vent à trous latéraux," Doctoral thesis, Université du Maine, France.

Murray, J. D. (**1993**). *Mathematical Biology* (Springer, Berlin).

Polack, J., Meynial, X., Kergomard, J., Cosnard, C., and Bruneau, M. (**1987**). "Reflection function of a plane sound wave in a cylindrical tube," Rev. Phys. Appl. **22**, 331–337.

Raman, C. V. (**1918**). "On the mechanical theory of vibrations of bowed strings," Indian Assoc. Cult. Sci. Bull. **15**, 1.

Schumacher, R. T. (**1981**). "*Ab initio* calculations of the oscillations of the clarinet," Acustica **48**, 72–85.

Valette, C., and Cuesta, C. (**1993**). *Mécanique de la Corde Vibrante* (Editions Hermés, Paris).

Woodhouse, J. (**1993**). "On the playability of violins. I. Reflection functions," Acustica **78**, 125–136.

Woodhouse, J., and Loach, A. R. (**1999**). "The torsional behavior of cello strings," Acta Acust. (Beijing) **85**, 734–740.

The design of bells with harmonic overtones

Neil McLachlan[a] and Behzad Keramati Nigjeh
Department of Aerospace Engineering, RMIT University, GPO Box 2476V, Melbourne 3001, Australia

Anton Hasell
Australian Bell Pty Ltd, Redesdale, Victoria 3444, Australia

(Received 12 December 2001; accepted for publication 21 March 2003)

Musical bells have had limited application due to the presence of inharmonic partials in the lower part of their acoustic spectra. A series of bells has been designed that contains up to seven partial frequencies in the harmonic series beginning at the fundamental frequency. This was achieved by choosing geometries for finite-element analysis models in which as many purely circumferential bending modes as possible occurred at frequencies below any mode with an axial ring node. The bell models were then fine tuned using gradient projection method shape optimization and the resulting profiles were cast in silicon bronze. A range of bell geometries and timbres is analyzed using psycho-acoustic models and is discussed in relation to European carillon bells. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1575748]

PACS numbers: 43.75.Kk, 43.40.Ey, 43.20.Ks [NHF]

I. INTRODUCTION

The pre-Christian use of the bell, particularly in Europe, appears predominantly to have been apotropaic (to scare off evil spirits). The earliest bells found in the west, from around 1200 BC, are pellet bells. These were worn on clothes and attached to horse harnessing to act as charms to protect the wearer from unnatural harm.[1] However, it was the reverence the early Christians held for the bell and its uses that ensured its continued rise in importance across Europe. Beginning with bell founding in monasteries, small handbells became transformed into stationary bells. The building of the great Gothic cathedrals raised the bells to greater heights from which bells of increasing size could broadcast their sounds to growing Christian communities.[1]

The first reported attempt to tune the partial frequencies of bells relates to Jacob van Eyck (1540–1657), who worked with François and Pieter Hemony on a commission to cast a carillon for the Wijnhuis tower in the town of Zutphen in 1643.[2] The carillon was considered markedly superior to all previous carillons. After the Hemony brothers died, the understanding of tuning of bells fell into disarray and was lost.[2] Bell tuning was recovered from the collaboration between bell enthusiast Canon Arthur B. Simpson, Rector of Fiddleworth, Sussex and the Taylors Bell-founders of Loughborough.[3] Simpson proposed a method of tuning, similar to the Hemony process, where the bell was cast oversize and lathed in particular segments of the waist or the lip of the bell to bring the partial tones into a harmonious series. Their first ''harmonic-tuned bells'' were hung at St. Paul's Church, Bedford, during 1896.[4]

The frequency ratios of the first five radiating partial tones in these Taylors' bells (and common to many tuned European bells since) is 1:2:2.4:3:4. Traditionally these partials are called the ''hum,'' ''fundamental,'' ''tierce,'' ''quint,'' and ''nominal.'' Thus, the European bell could be tuned to a series of partials that included the first four harmonics. However, the presence of the tierce is likely to create complex acoustical percepts of pitch and dissonance that change in time as partial frequencies decay at varying rates. In their discussion of the pitch perception of European bells, Schouten and 't Hart[5] remark, ''If one concentrates on these partials when listening to a ringing bell, it is a remarkable experience to hear them with every strike. The minor third is quite conspicuous and it diminishes only slightly in loudness between strikes.''

The strike tone of the European bell is usually reported as the fundamental and has been shown to be largely influenced by the frequency of the 5, 6, and 7th overtones when subjects attempt to nominate a pitch for the bell.[6] This synthetic percept is an example of a virtual pitch or complex tone sensation arising from a different listening attitude to concentrating on individual partials. It will vary depending on the type of clapper and on how hard the bell is struck.

The pitch of bells can also be described by the salience and multiplicity of simultaneously audible tones and pitch sensations at various times in the bell sound. Mathematical models of pitch multiplicity have been developed and compared to human subjective responses to synthesized bell timbres and a range of musical chords and complex harmonic tones. The calculated and reported values of multiplicity for these synthesized bell sounds were found to be higher than for complex harmonic tones with comparable fundamental tone frequencies.[7] These models first calculate the audibility of pure tones (levels above the masking threshold) and any virtual pitch tones that may then arise. The multiplicity is calculated by first normalizing the audibility of these tones to the most audible tone, summing the normalized tone audibility, and then dividing by the audibility of the most audible tone. A factor determined from results of human subjective tests can then be used to adjust for the listening attitude, where analytical listening favors hearing individual partials. The tone salience of individual pure-tone or complex tone percepts is proportional to their calculated audibility.[7]

[a]Electronic mail: neil@ausbell.com

Bells do not naturally produce harmonic overtones because, unlike air columns and strings that vibrate predominantly in one dimension only, bells vibrate flexurally in three dimensions. Flexural vibrations are much more difficult to describe analytically than longitudinal vibrations, and it is common to use numerical methods such as finite-element analysis (FEA) to predict the behavior of bells and gongs.[8,9] A number of major third carillons (where the partial is tuned to 2.5 times the hum tone frequency instead of 2.4) have been designed using finite-element analysis in the Netherlands.[10]

II. OPTIMIZATION METHOD

Given the ability to predict natural frequencies from computer models of bells, gradient projection method shape optimization can be applied to the models to adapt wall profiles and, where possible, arrive at specific tuning ratios of partial frequencies. Predeveloped, commercial FINITE-ELEMENT (FE) software was employed to implement a classic linear finite-element method to numerically determine the natural frequencies of the model.[11] The ''solid property'' elements used for the modeling consisted of *tetra* and *penta* elements in which the nodes' coordinates determine some subsequent shape parameters such as thickness along the bell length. The accuracy of FEA-predicted frequencies for the lower frequencies of bells has been shown to be around 1% of measured results[9] when working with two-dimensional FE models. It is reasonable to expect greater accuracy from three-dimension solid property models.

A gradient projection method[12] was also utilized in this software in order to optimize the objective parameters by changing the coordinates of the FE nodes. This software is designed to compute the nodal vector sensitivities to the objective parameters (such as modal natural frequencies) as a function of the FE nodes' coordinates. The user may select a zone of active nodes with coordinates that will vary during optimization. The sensitivity is calculated from differences in the objective parameter after displacement of each active finite-element node. The process of optimization then iterates towards a target in accordance with geometric constraints that preserve shape parameters of the model such as symmetry about the vertical axis.

To achieve the final goal of optimization (in this case, tuning the modal frequencies of the first seven modes to the harmonic series), the optimization process would ideally be performed separately on each vibrational mode. Behavioral constraints can be applied to limit the allowable changes of frequency of particular modes so that optimization processes on other modes do not change them. These behavioral constraints are met by computing the nodal vector sensitivities to the constraint parameters as a function of the nodes' coordinates. The final displacement vector for each active FE node is calculated from sensitivities to both constraint and optimization parameters during each iteration of the optimization process. The optimization process stops if progress toward the optimization target cannot be achieved without altering the constrained parameters beyond a given tolerance.

The user sets the step size used in the first iteration of the optimization process. It is also possible to define a reduction rate for the step size in the following iterations in order to prevent the optimization from overshooting the target and then alternating on either side of it. The initial step size and its reduction rate were carefully selected based on experience with similar models.

III. SHAPE OPTIMIZATION OF BELLS

It was quickly observed that the behaviors of many bending modes are strongly correlated with respect to changes in the geometry of the model. Therefore, it is often not possible to raise or lower any given partial frequency using gradient projection method optimization applied to positions of FE nodes without affecting other partial frequencies. Constraining all the modes except the mode being optimized restricts the optimization process, and in most cases the objective value cannot be achieved.

A number of strategies may be employed to overcome this problem. The maximum vibrational amplitudes of the bending modes to be tuned may occur in different locations along the length of the bell, and therefore the magnitude of their sensitivity vectors may also vary. It may be possible for the constrained modes to remain relatively unaffected by the optimization process if a careful selection of the active nodal zones is made on the basis of the relative positions of vibrational maxima.

A study of the behavior of modes for changes in model geometry could enable the user to identify which mode types are correlated for certain geometrical changes. By careful selection of the sequence of modes to be optimized, and the constraint sets, groups of highly correlated modes could be moved simultaneously toward a set of target frequencies. Furthermore, the deviation tolerance of the constrained modes' frequencies may be increased to allow them to change more during the optimization. This may be sufficient

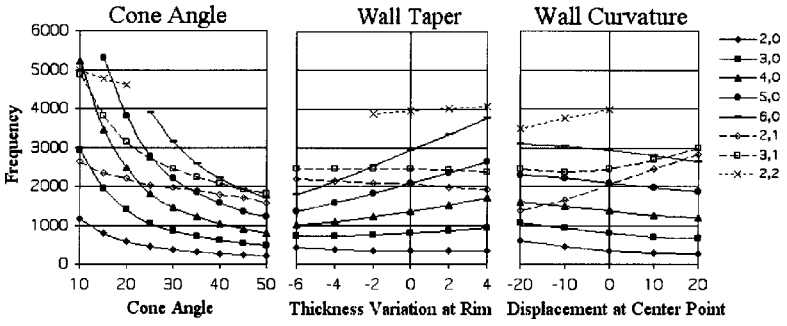


FIG. 1. FEA-predicted frequencies for varying the cone angle, wall taper, and wall shape of freely vibrating capped cones.

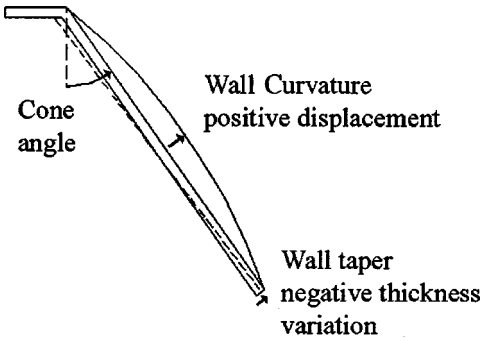


FIG. 2. The geometric parameters used to produce the data shown in Fig. 1.

to achieve a set of optimization targets within an acceptable error.

In order to better understand these correlated behaviors a series of experiments was conducted in which only one geometric parameter of an FE model was varied, and the resultant changes to the predicted partial frequencies of various modes were tabulated and plotted. The effects of geometric parameters such as wall curvature and taper, and cone angle for cylinders and cones with closed ends were investigated for models able to vibrate freely. The results for varying cone angles have been reported in a previous paper.[13] Results for purely circumferential modes (modes which produce only nodal lines in plane with the axis of symmetry of the bell), and mixed modes (producing both nodal lines and rings) are reported below.

Wave number pairs (m,n) are used to describe bending modes such that the first number refers to the number of nodal lines in plane with the axis of symmetry (i.e., circumferential wave number), and the second number refers to the number of nodal rings perpendicular to the cylinder length (i.e., axial wave number). The data from the experiments referred to above could then be used to select a model geometry in which the partial frequencies were reasonably close to the harmonic series and, given an understanding of the behavior of the modes, likely to be able to be tuned correctly using gradient projection method shape optimization. Figure 1 shows plots of FE predicted frequencies for varying cone angle, wall taper, and wall curvature of a capped cone of 200-mm length, 220-mm top circumference, and 10-mm wall thickness. These parameters are defined in Fig. 2. Silicon bronze (95% copper, 1% manganese, and 4% silicon) and its material properties were used for all the models and cast bells described in this paper. The mass density and Young's modulus of silicon bronze are 8.4×10^3 kg/m^3, and 9.4×10^{10} N/m^2, respectively.

From this figure it can be seen that the frequencies of the m,0 modes remain approximately equally spaced as the range of their frequencies change with these geometric changes. At a cone angle of 10° the modes are in the order 2,0; 2,1; 3,0; 3,1; and 4,0, which is the order found in European bells. Mixed modes decrease in frequency more slowly than purely circumferential modes with increasing cone angles and increasing wall tapers toward thinner rims. Increasing curvature away from the axis of symmetry (concavity) increases the frequency of the mixed modes and decreases the frequency of purely circumferential modes.

Figure 1 reveals that the modes with the same number of axial nodes (n) are generally correlated in their frequency responses to changing geometry. This is most evident in the behavior of the m,0 modes described above, but is also clear from the data for the m,1 modes shown in Fig. 1. In setting a frequency optimization target for any given vibrational mode, it was clear that the correlated modes would behave in a similar fashion. For example, it was not possible to substantially change the frequency of the 3,0 mode while constraining the 2,0 and 4,0 modes. As can be seen from Fig. 1 the presence of an m,1 mode between two m,0 modes creates a very uneven distribution of frequencies. Given these modal behaviors, it was realized that to tune the partials to the harmonic series using gradient projection method optimization it would be necessary to first separate the mixed modes from the m,0 modes. The simplest geometrical solution to achieve this was high cone angles and a tapered wall thickness to a thinner rim.

Table I shows the results of optimization starting from a simple truncated cone with a tapered wall and resulting in a harmonic bell with the profile shown in Fig. 3. The entries shown in column 1 of Table I refer to the three geometric models involved in the optimization process, and column 2 refers to the optimization settings required to arrive at that geometry. For example, *down1 constrain3* means drop the frequency of mode 1 in geometry 1 to the target frequency shown for geometry 2, while constraining the frequency of mode 3. From the results of the first stage of the optimization, it can be observed that expanding the frequency range between the 2,0 and 4,0 modes (by dropping the frequency of mode 1 while constraining mode 3) causes a similar expansion in the frequency range between the 4,0 and 7,0 modes. Since the 4,0 mode was constrained, this raised the

TABLE I. Sequential optimization results using "RESHAPE" for the tuning of a 7-partial, harmonic bell starting from a truncated cone with tapering wall thickness.

Step	Mode no.	1	2		3		4		5		6		7	
	Mode type	2,0	3,0		4,0		5,0		6,0		7,0		2,1	
		Freq	Freq	Ratio	Freq	Ratio	Freq	Ratio	Freq	Ratio	Freq	Ratio	Freq	Ratio
1	Initial tapered cone	188	370	1.97	540	2.87	704	3.74	871	4.63	1060	5.64	1260	6.7
	% error			−1.6		−4.3		−6.4		−7.3		−6		−4.3
2	Down 1 constrain 3	180	357	1.98	538	2.99	727	4.04	925	5.14	1141	6.34	1256	6.98
	% error			−0.8		−0.4		0.97		2.78		5.65		−0.3
3	Down 6 constrain 1,2,3,7	181	359	1.98	535	2.96	711	3.93	887	4.9	1079	5.96	1255	6.93
	% error			−0.8		−1.5		−1.8		−2		−0.6		−0.9

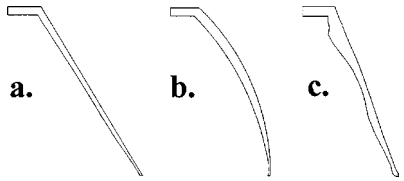


FIG. 3. (a) FE profile of a 7-partial, conical, harmonic bell; (b) 7-partial, concave, harmonic bell; and (c) FE profile of a 5-partial, conical, harmonic bell.

frequency of the higher frequency m,0 modes. The 2,1 mode, however, was only slightly affected. A slight correction of the higher m,0 mode frequencies was then possible in the final stage of the optimization.

A maximum allowable tuning error for the overtones was set at 2% from the harmonic series. Human frequency discrimination of pure tones of 0.5-s length and between 125 and 1000 Hz varies between 0.2% and 1.7% depending on the frequency and the method used to determine the just-noticeable difference.[14,15] For pure tones 10 dB above a partially masking broadband noise, the just-noticeable frequency difference doubles and continues to increase for louder masking noise.[16] Since all the partials heard in a bell are partially masking each other, a tuning error of less than 2% for the upper partials was considered reasonable. The probability that a mistuned harmonic is detected as being different from the harmonic series by automatic neural processes has been shown to increase substantially for mistunings greater than around 4% as determined by event related potentials (ERP) in EEG recordings. Subjective responses and ERP attributed to active decision-making processes for hearing the mistuned harmonic as a separate auditory entity increase substantially from a probability of around 10% for mistunings of the second harmonic greater than 2%. The probability of segregating the mistuned harmonic generally decreased for higher harmonics.[17]

IV. DESIGN VARIANTS

The profile shown in Fig. 3(a) (profile 1) will produce a fundamental frequency of 220 Hz for a bell with a diameter of 395 mm, height of 248 mm, and mass of 19 kg. European tuned bells with the same fundamental frequency (hum tone) and a strike tone of 440 Hz are about twice the diameter and weigh about 250 kg or more depending on the profile. As a consequence, European bells of the same fundamental frequency are able to produce much greater loudness than bells of this design. To increase the applicability of harmonic bells to a range of musical contexts, design variants were developed that were physically larger for the same fundamental tone.

Figure 1 reveals that creating models with increasing concavity (shown as positive curvature displacement in Fig. 2) increases the frequency of the mixed modes and decreases the frequency of purely circumferential modes. This enabled concave harmonic bells with seven tuned partials to be tuned by this method of shape optimization. These bells have greater mass and surface area than the equivalent conical bell and have been cast with a fundamental frequency of 75 Hz and a mass of 1150 kg. At a fundamental frequency of 220 Hz they will weigh 44 kg. Close inspection of Fig. 3(b) (profile 2) will reveal a small cylindrical section that was added to the bottom of the bell profile to further raise the frequency of the mixed modes relative to the circumferential modes prior to beginning the shape optimization.

The rims of the bells described above are thin relative to the rest of the bell. If these bells were scaled down in size to produce a fundamental frequency of greater than 440 Hz, the rims would be so thin that the bells could be susceptible to damage when struck (frequency is inversely proportional to size if scaled in all three dimensions). Finite-element studies indicated that increasing wall thickness increases the frequency of all bending modes. The frequencies of mixed modes do not increase as quickly as purely circumferential modes in capped cones, and so fewer purely circumferential modes have frequencies below the frequency of the first mixed mode and can be tuned by this method of shape optimization to the harmonic series.

A thicker conical bell was designed and cast with six partial frequencies in the harmonic series to compare with the conical bell with seven partial frequencies described above. This bell had a fundamental frequency of 215 Hz but was 550 mm in diameter, 305 mm in height and weighed 29 kg. Apart from the increased wall thickness, the bell had a very similar profile to the conical bell shown in Fig. 3(a) and is described as profile 1a in Table II.

Since the partial frequencies being tuned are predominantly due to circumferential bending modes, reducing the circumference of the bell at the rim by decreasing the cone angle further increases the frequency of the fundamental without decreasing the length and thickness of the wall (see Figs. 1 and 2). However, the frequencies of the mixed modes do not increase as quickly as the purely circumferential modes, and so fewer partials can then be tuned to the harmonic series by gradient projection method optimization applied to FE nodal positions. Maximizing the number of par-

TABLE II. Frequencies measured by peak detection of FFT spectra.

		Measured frequencies of bells described in figures						
Profile	Mass (kg) @ 220 Hz	Freq 1 (Hz)	Freq 2 (Hz)	Freq 3 (Hz)	Freq 4 (Hz)	Freq 5 (Hz)	Freq 6 (Hz)	Freq 7 (Hz)
1.	19	207	419	613	814	1018	1249	1434
1a.	29	345	692	1033	1362	1709	2062	NA
2.	46	188	380	567	746	924	1116	1313
3.	83	232	465	699	944	1176	NA	NA

tials tuned to the harmonic series resulted in much greater wall tapers with increased wall thickness near the cap of the bell.

Conical harmonic bells with 5, 4, and 3 partial frequencies tuned to the harmonic series have been designed and cast to enable bells with fundamental frequencies as high as 1720 Hz (the note E5) to be produced. These bells have cone angles as low as 13.5 degrees. Figure 3(c) shows the profile of a 5-partial harmonic bell (profile 3). This profile produces a fundamental of 232 Hz for a bell of 520-mm diameter, 460-mm height, and weighing 83 kg. Careful inspection of the profile will reveal that the outside wall is not a straight line as in other conical bells, as the FE nodes comprising this wall were also allowed to move in the optimization process to maximize the tuning possibilities. This profile is expected to be appropriate for church and carillon towers (except for full-circle ringing) as well as for higher frequency orchestral bells.

V. ANALYSIS OF CAST BELLS

Tuning errors were encountered during the manufacture of many bells. These were predominantly ascribed to geometric inaccuracies or variation in the metal properties created in the casting of the bell. It was found to be possible to fine-tune these bells on a lathe after casting, by using shape optimization experiments as a guide as to where to remove material. The frequency of the lower frequency modes could be decreased by reducing the wall thickness near the cap, and increased by reducing wall thickness near the rim. Lower frequency modes could be decreased without affecting higher frequency modes; however, increasing them by thinning the wall near the rim caused the frequency of the higher frequency modes to decrease. All modes increase in frequency if the bell is shortened in length. If retuning was found to be necessary it sometimes resulted in the highest frequency tuned mode falling outside the 2% error margin.

The acoustic spectra of cast harmonic bells with profiles 1a to 3 are shown in Fig. 4. The actual frequencies shown in the spectra of the bells are given in Table II along with the frequencies measured for a bell cast with profile 1. All the tuned frequencies for bells with profiles 1 to 3 are within 2% of the harmonic series. Table II also includes the mass for a bell of each profile when scaled to produce a fundamental frequency of 220 Hz.

The acoustic spectra were produced from recordings taken at 1 m from the bell surface perpendicular to the axis of symmetry of the bells. Bell 1 was struck near the rim by a steel hammer with a hard nylon head of 0.5-kg mass, bell 2 by a steel hammer with a hard nylon head of about 1.5-kg mass, and bell 3 by a spherical cast-iron clapper of about 5-kg mass. These mallets were chosen as appropriate for producing a tone balance predominantly consisting of the full range of tuned partials. The mallet velocity was such as to produce an ''A-weighted'' sound-pressure level of around 85 dB (fast response) in the early part of the sound. The FFT was produced at about 50 ms after the sound onset by using a Hamming window of 4096 samples for a sample rate of 44.1 kHz.

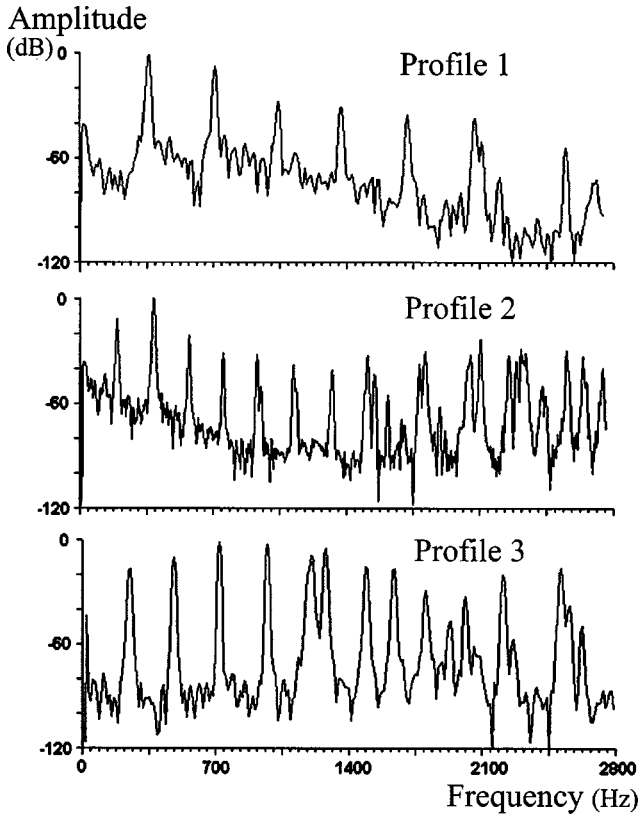


FIG. 4. Acoustic spectra recorded 50-ms. after the sound onset for bells described in Table II.

The strategy of increasing the loudness of conical bells by increasing their mass at the expense of the number of tuned partials has been evaluated as follows. The difference in the perceived timbre of conical bells with 6 or 7 partials was only slight, but the larger and thicker bell was able to produce about 10 dB(A) greater peak sound-pressure level (fast response) for a range of strike velocities between 1 and 2.5 m/s with a typical tubular bell mallet (leather-clad hardwood) or a steel hammer with a hard nylon head and mass of about 0.5 kg. Critical bandwidths increase in proportion to frequency for frequencies above 500 Hz.[18] Therefore, for higher harmonic numbers, and higher fundamental frequencies, the more partial frequencies occur in the same critical band and the greater partial masking will occur between them. Comparing the probability of reporting mistuned harmonics as separate auditory entities from otherwise harmonic complex sounds with fundamental frequencies of 200 and 400 Hz has shown the diminishing importance of higher partials to the perception of pitch. The probability of segregating the sixth harmonic mistuned upwards by 5.3% was reduced from about 80% to about 30% when the fundamental frequency was doubled.[17]

Psycho-acoustic models have been employed to evaluate the bells prior to undertaking more extensive subjective tests relevant to various musical contexts. If the high-frequency untuned partials do contribute more than a general sharpness percept in the early portion of the sound, they will increase the pitch multiplicity of the bell. This was measured using Parncutt's algorithm[7] applied to FFT spectra of the bells calculated using a Hamming window of 2048 samples at a

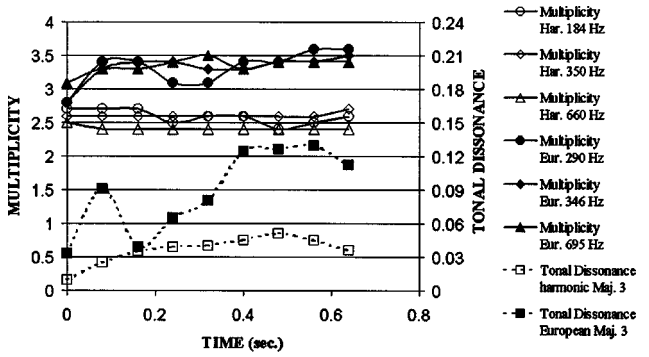


FIG. 5. Comparison of calculated pitch multiplicity and tonal dissonance for a range of harmonic bells and European carillon bells described by the frequencies of their first partial. The harmonic bells at 188 and 345 Hz are described in Table II and Fig. 4 with profiles 2 and 1a, respectively.

sample rate of 12.8 kHz.[19] Pitch multiplicity can be as high as 3 for harmonic complex tones due to the audibility of individual harmonics and the presence of multiple virtual pitches or subharmonic pitch percepts. Figure 5 shows pitch multiplicity calculated over 80-ms intervals from the sound onset for a range of harmonic bells including the bells with profiles 1a and 2 described in Table II, and three European carillon bells with comparable frequencies. The harmonic bell with a fundamental at 660 Hz is a conical 5-partial bell with a profile similar to profile 3. The pitch multiplicity of all three harmonic bells remains around 2.5 for the first 640-ms of the sound, and generally decreases as pitch increases, despite the number of tuned partials decreasing from 7 to 5. The calculated multiplicity over the same time period for the larger 5-partial bell described in Table II is shown in Fig. 6,

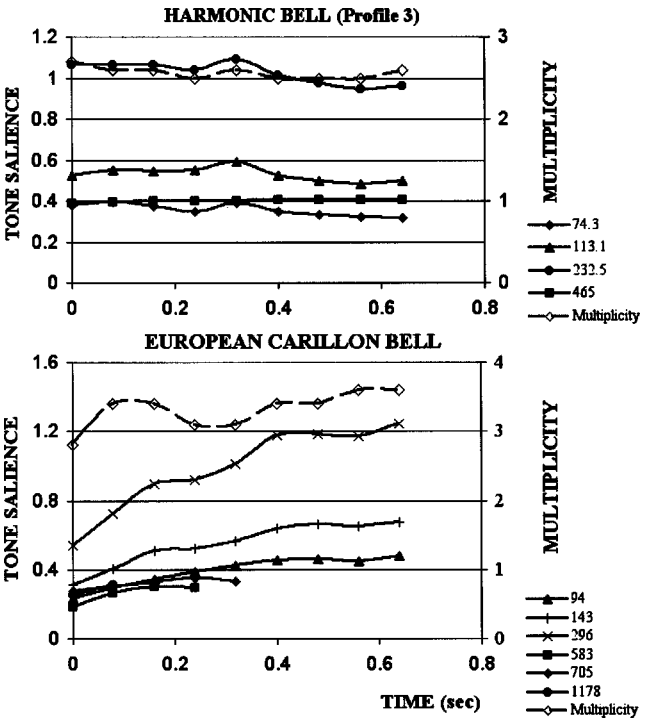


FIG. 6. Comparison of calculated tone salience and pitch multiplicity for a lower pitched 5-tone harmonic bell (profile 3 described in Table II) with a European carillon bell.

and is similar to the harmonic bells described in Fig. 5. This suggests that there is little improvement in the pitch clarity gained by tuning partials beyond the fifth partial for harmonic bells of this pitch.

All the harmonic bells have much lower calculated pitch multiplicity than the European bells (recorded in a similar manner *in situ* at the War Memorial Carillon, Loughborough, UK). The European bells' calculated pitch multiplicity (shown in Fig. 5) increases after the first 80-ms interval, as masking effects in the sharp attack portion of the sound diminish and more partial frequencies become audible. Figure 6 shows the calculated pitch multiplicity and tone saliences over the same time period for the lowest pitch European bell. Tone salience was calculated using Terhart's algorithm.[20] The audibility of the tierce at 705 Hz, the fundamental at 583 Hz, and the nominal at 1178 Hz can be seen to contribute to the calculated pitch multiplicity for the first 300 ms. Thereafter, the hum at 296 Hz and its subharmonics become the principal contributors to the pitch multiplicity.

These results suggest that the perceptual prominence of a single strike tone in European bells lasts for less than 80 ms when the bell is struck with moderate force in the manner of a carillon playing. While the attack portion of musical sounds influences judgments about the rest of the sound, it is likely that large changes in the perceived timbre occurring after the attack portion, such as those shown above for European bells, will also be salient. Figure 6 also shows comparative tone salience and multiplicity results for the heavy 5-partial, harmonic bell described in Table II (profile 3). The contributions to the calculated multiplicity of the second partial at 465 Hz, and the first partial at 232 Hz and its subharmonics remain consistent throughout the first 640 ms of the sound.

Figure 5 also shows the tonal dissonance for major third dyads of harmonic bells and European carillon bells calculated over 80-ms intervals using the algorithm developed by Hutchinson and Knopoff[21] applied to FFT produced as described above. The 290- and 346-Hz European bell sounds were used with the 346-Hz sound pitch shifted up one semitone using a digital sampler. Two harmonic bells of the same fundamental frequencies were also used. Tonal dissonance is calculated from the sum of dissonances between partial tones. This is derived by multiplying the measured amplitudes and frequency differences between partials by roughness factors associated with critical bandwidths determined for each frequency difference. The sum is then normalized by dividing by the total power of these partials. The dissonance of major third dyads calculated for the European bells is consistently greater than for the harmonic bells and varies in a similar fashion to the changing multiplicity of the bell sounds used in the calculation.

Another factor influencing the loudness and perceptibility of high-frequency partials is the internal damping of the bronze. When struck by a hard wooden beater, silicon bronze bells with fundamental frequencies above 1200 Hz produce a fundamental frequency with a sound-pressure level more than 50 dB greater than any of the higher partials when measured at 1-m distance perpendicular to the axis of symmetry. The fundamental will dominate in the perception of pitch in

these bells, and possibly mask all other spectral components of the sound. It is therefore appropriate to design bells with as few as three tuned partials as long as they are high pitched. The material damping in tin–bronze bells (for alloys of 10%–22% tin) is less than silicon bronze bells, and so they sustain the resonance of higher partials for longer times. However, tin–bronze is also more susceptible to cracking, especially with the thin rims necessary to produce harmonic bells, and so only one prototype harmonic bell has been cast in tin–bronze to date. European-style handbells are made in tin–bronze and generally only have two tuned partials, although many more are distinctly audible.[22]

VI. CONCLUSION

The combination of gradient projection method shape optimization of FE nodal positions with the FE frequency modal analysis data for variations of a series of geometric parameters enabled bells with harmonically tuned partials to be designed. The design process involved the careful choice of a starting geometry for shape optimization based on knowledge of the frequencies of modes and the correlated behavior of various lower frequency mode types for cones and cylinders with geometries in the range of interest. This process often required the designer to make changes to the initial geometry of the model based on the broadly determined parametric data after a gradient projection method optimization failed to reach optimization targets.

Harmonic bells have been designed and manufactured with less than 2% error for up to the first seven partial frequencies. These bells are expected to expand the musical applications of bells due to reduced complexity of pitch percepts in the sound of the bells and partials likely to create dissonance in musical chords. The effects of the tuning errors of the upper partials will be on the clarity of the perceived pitch when the bell is heard in isolation, and the roughness due to beating partials when played in musical chords. In the bells manufactured to date[23] these affects are very small, but will require more research to properly quantify.

Harmonic bells have been used in a number of musical contexts. Harmonic bells ranging in mass from over 1 ton to less than 1 kg are included in a permanent outdoor installation of electro-mechanically struck bells in Melbourne, Australia. These bells are struck by a variety of mallet materials ranging from steel pins for the smaller bells to metal mallets with hard nylon striking surfaces for the larger bells. They are controlled by MIDI programs and can be played with a range of mallet velocities. This installation is not intended to be as loud as a chime of European bells, but nevertheless can be readily heard from over 100-m distance. Over 2000 harmonic handbells and a two-octave set of larger harmonic bells for the Melbourne Symphony Orchestra were produced for performances to celebrate the centenary of Australia's federation in 2001.[23]

ACKNOWLEDGMENTS

The authors would like to thank the State of Victoria and the Melbourne International Festival of the Arts for supporting this work.

[1] P. Price, *Bells and Man* (Oxford University Press, Oxford, 1983).
[2] A. Lehr, *The Art of the Carillon in the Low Countries* (Iannoo, 1991).
[3] T. S. Jennings, *Master of My Art—The Taylor Bellfoundries 1782–1987* (Taylors Bellfoundry, Loughborough, 1987).
[4] A. B. Simpson, "Why Bells Sound Out of Tune," Pall Mall Magazine, October 1895.
[5] J. F. Schouten and J. 't Hart, "The Strike Note of Bells," in *Acoustics of Bells* (Van Nostrand Reinhold, New York, 1984), translated by 't Hart from original publication "De Slagtoon Van Klokken," Neth. Acoust. Soc. Pub. No. 7, 8–19 (1965).
[6] N. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, New York, 1990), Chap. 21.
[7] R. Parncutt, *Harmony: A Psychoacoustical Approach* (Springer, Berlin, 1989), Chap. 5, pp. 85–105.
[8] N. McLachlan, "Finite Element Analysis and Gong Acoustics," Acoust. Australia **25**(3), 103–107 (1997).
[9] R. Perrin and T. Chanley, "The Normal Modes of the Modern English Church Bell," J. Sound Vib. **90**, 29–49 (1983).
[10] A. Schoofs, F. Van Aspern, P. Maas, and A. Lehr, "A Carillon of Major-Third Bells. I. Computation of Bell Profiles using Structural Optimization," Music Percept. **4**(3), 245–254 (1987).
[11] J. Tomas, RESHAPE *User's Manual, Version 98a* (Advea Engineering, Melbourne, 2000).
[12] R. L. Fox, *Optimization Methods For Engineering Design* (Addison-Wesley, Reading, MA, 1971), pp. 196–222.
[13] N. McLachlan and B. K. Nigjeh, "Investigations of the Vibrational Behaviour of Cylindrical Forms with Application to Musical Instrument Design," in *EMAC 2000 Proceedings* (The Institute of Engineers, Australia, 2000), pp. 215–218.
[14] H. Fastl and A. Hesse, "Frequency Discrimination for Pure Tones at Short Durations," Acustica **56**, 41–47 (1984).
[15] W. M. Hartmann, "Pitch, Periodicity, and Auditory Organization," J. Acoust. Soc. Am. **100**(6), 3491–3502 (1996).
[16] E. Zwicker and H. Fastl, *Psycho-acoustics: Facts and Models*, 2nd ed. (Springer, New York, 1999), Chap. 7, p. 193.
[17] C. Alain, S. R. Arnott, and T. W. Picton, "Bottom-Up and Top-Down Influences on Auditory Scene Analysis: Evidence From Event-Related Brain Potentials," J. Exp. Psychol. Hum. Percept. Perform. **27**(5), 1072–1089 (2001).
[18] E. Zwicker and H. Fastl, *Psycho-acoustics: Facts and Models*, 2nd ed. (Springer, New York, 1999), Chap. 6, p. 161.
[19] N. McLachlan and D. Cabrera, "Calculated Pitch Perceptions for New Musical Bell Designs," in *ICMPC7 Conference Proceedings* (University of Western Sydney, Sydney) (to be published).
[20] E. Terhardt, G. Stoll, and M. Seewann, "Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals," J. Acoust. Soc. Am. **71**(3), 679–688 (1982).
[21] W. Hutchinson and L. Knopoff, "The Acoustical Component of Western Consonance," Interface (Netherlands) **7**, 1–29 (1978).
[22] T. D. Rossing and H. J. Sathoff, "Modes of Vibration and Sound Radiation from Tuned Handbells," J. Acoust. Soc. Am. **68**(6), 1600–1607 (1980).
[23] To see and hear a wide range of harmonic and other bells see the Australian Bell website, www.ausbell.com

Acoustic communication in two freshwater gobies: Ambient noise and short-range propagation in shallow streams[a)]

M. Lugli[b)]
Dipartimento di Biologia Evolutiva e Funzionale, Università di Parma, Parma, Italy

M. L. Fine
Department of Biology, Virginia Commonwealth University, Richmond, Virginia 23284-2012

(Received 14 December 2002; revised 22 March 2003; accepted 24 March 2003)

Noise is an important theoretical constraint on the evolution of signal form and sensory performance. In order to determine environmental constraints on the communication of two freshwater gobies *Padogobius martensii* and *Gobius nigricans*, numerous noise spectra were measured from quiet areas and ones adjacent to waterfalls and rapids in two shallow stony streams. Propagation of goby sounds and waterfall noise was also measured. A quiet window around 100 Hz is present in many noise spectra from noisy locations. The window lies between two noise sources, a low-frequency one attributed to turbulence, and a high-frequency one (200–500 Hz) attributed to bubble noise from water breaking the surface. Ambient noise from a waterfall (frequencies below 1 kHz) attenuates as much as 30 dB between 1 and 2 m, after which values are variable without further attenuation (i.e., buried in the noise floor). Similarly, courtship sounds of *P. martensii* attenuate as much as 30 dB between 5 and 50 cm. Since gobies are known to court in noisy as well as quiet locations in these streams, their acoustic communication system (sounds and auditory system) must be able to cope with short-range propagation dictated by shallow depths and ambient noise in noisy locations. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1577561]

PACS numbers: 43.80.Ka, 43.80.Ev, 43.30.Xm [WA]

I. INTRODUCTION

Sound attenuation (spreading loss and absorption), degradation, and ambient noise act as environmental constraints on acoustic communication (Wiley and Richards, 1982). For the emitter these factors are thought to be important for the evolution of animal vocalizations, and for the receiver, they affect the detection and recognition of sounds. Therefore, these factors are important for the design of the auditory system. Evidence for the effects of the environmental factors on acoustic signals and receptor systems has been provided for many terrestrial species, especially among birds and mammals (reviewed in Bradbury and Veherencamp, 1998).

Acoustic communication is also affected by the underwater environment. For instance, the short travel time of sound in water and the air–water and water–bottom boundaries account for the strongly frequency-dependent propagation and high degradation typical of underwater acoustic signals (e.g., Hawkins and Myrberg, 1983). These effects are pronounced in coastal waters in the sea and in shallow waters of rivers, lakes, and ponds where most sonic teleosts are found. Acoustic characteristics of noise in the deep sea are well known (Knudsen *et al.*, 1948; Wenz, 1962; reviewed in Urik, 1983), but little work has been devoted to shallow environments (Hawkins and Johnstone, 1978; Myrberg, 1980; Fine and Lenhardt, 1983; Forrest *et al.*, 1993; Mann and Lobel, 1997).

Teleosts have the most diverse sound-producing mechanisms, hearing abilities, and live in a wider array of environments than other vertebrate groups. However, because of low accessibility for experimental investigation and complex acoustics of the aquatic environment, the relationships between the environmental constraints and sound communication in fishes are less understood than in land vertebrates. Rogers and Cox (1988) suggested that high noise levels at low frequencies (<1 kHz) in many shallow-water environments should favor fishes that preferentially evolve sensitivity to high frequencies. Yet, many teleosts from shallow habitats produce sounds and have best hearing at frequencies well below 1 kHz (e.g., Fine *et al.*, 1977; Myrberg, 1981; Hawkins and Myrberg, 1983; Fay and Popper, 1999). The presence of physiological constraints (swimbladder resonance) or short-range communication are suggested explanations for the paradox (Fine and Lenhardt, 1983; Bradbury and Veherencamp, 1998).

Two gobies (*Padogobius martensii*, *Gobius nigricans*) live in streams and small stony rivers (Lugli *et al.*, 1992; Gandolfi and Tongiorgi, 1974) characterized by low water depths (<1 m) and high levels of low-frequency background noise from water turbulence and small waterfalls. These habitats offer an excellent opportunity to study the role of ambient noise and other environmental constraints on fish sound communication. In this paper we characterize ambient noise levels in various locations (quiet to noisy) in two such shallow streams (Stream Stirone, River Serchio). We also measure the propagation of noise downstream from a waterfall and the propagation of *P. martensii* sounds in a quiet

[a)]Part of this work was presented in "Environmental constraints on the acoustic communication system of stream gobies," Extended Abstract in Bioacoustics, special issue of the symposium: Fish Bioacoustics: Sensory Biology, Behavior, and Practical Applications, Evanston, IL, 30 May–2 June 2001.

[b)]Electronic mail: lugli@biol.unipr.it

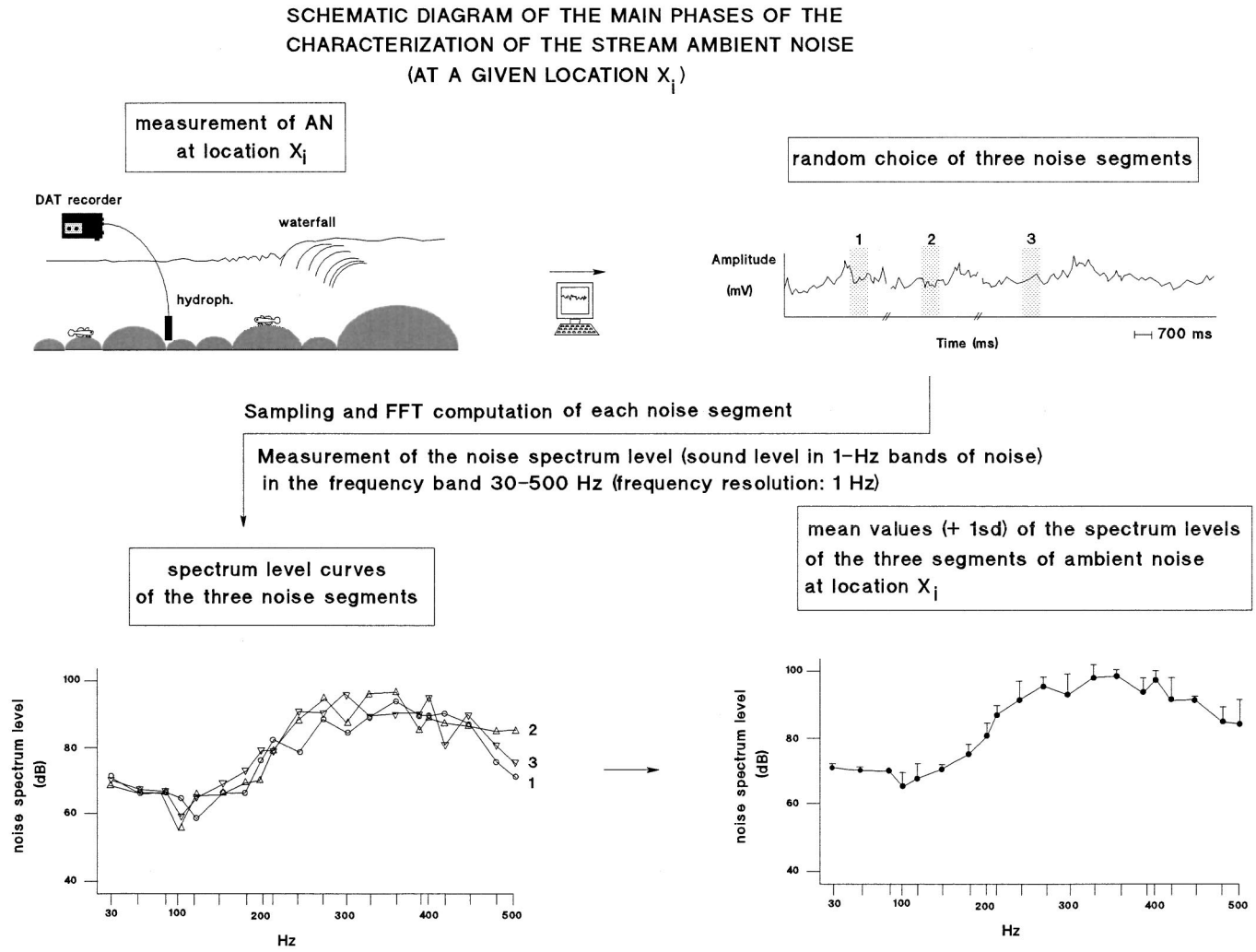


FIG. 1. Schematic diagram and main phases of the characterization of the ambient noise in the stream and in the laboratory. AN = ambient noise.

location of Stream Stirone. The results indicate severe constraints on acoustic communication both because of high noise levels in some subhabitats and short-range propagation of sound in shallow areas utilized by gobies for spawning. A window in the noise around 100 Hz coincides with the most sensitive hearing and the peak frequency of the sound spectrum in these fishes (Lugli *et al.*, 2003).

II. EXPERIMENTAL PROCEDURE

The study species come from two separate freshwater systems: *P. martensii* in Stream Stirone (a small hill stream located about 40 km west of Parma, Northern Italy), and *G. nigricans* in River Serchio (a small stony river located 2 km north of Lucca, Tuscany, Central Italy). The study site in Stream Stirone is 2.5 km long, with a width from 2 to 15 m, water depth usually <30 cm, and average gradient of 1.2 cm/m (Lugli *et al.*, 1992). The bottom consists mainly of flat stones and small areas of coarse gravel. Features of the stream vary greatly from place to place—due to changes in water current, depth, and bottom topography. The site in River Serchio is 14 km long, with a maximum width >50 m and a water depth up to 2 m. This river has a higher water discharge, a wider stream bed, and higher water depths than Stream Stirone. The bottom consists of stones and coarse gravel, but the stones are clearly larger and more rounded than in Stream Stirone. Lugli *et al.* (1992) showed that the distribution of breeding individuals on the bottom is unaffected by hydrological parameters of the stream, such as water depth, current speed, or distance from the stream's banks.

A. Ambient noise measurements

Ambient noise (AN) was measured at various locations where nesting gobies were found (Fig. 1). Locations included quiet areas and sites within 4 m of small waterfalls, rapids, and other places where the water surface breaks because of the presence of a big stone producing an overfall downstream. Sites were not picked randomly but were chosen to favor sites likely to have elevated noise levels (Fig. 1).

AN was measured with a preamplified pressure-sensitive hydrophone (ITC 8073, sensitivity: -167 *re*: 1 V/μPa, frequency response: ± 1.5 dB from 20 to 2000 Hz) placed on the bottom. The hydrophone was connected to a portable DAT recorder (Casio DA-7, sampling rate: 48 kHz). A single recording was made at each location for approximately 1 minute. AN measurements were made at 23 locations, of which 13 were from noisy areas, in Stream Stirone, and 16

locations, of which 10 where from noisy areas, in River Serchio.

Noise recordings were stored on a PC (sampling rate 5000 Hz), and analyzed using the AVISOFT software package for sound analysis. All recordings were low-pass filtered at 1 kHz to examine the low-frequency spectrum of the stream. For quantitative and statistical purposes, they were later bandpass filtered (30–500 Hz) to focus on frequencies important for goby acoustic communication. The AN spectrum was determined from three noise segments of approximately 700 ms that were randomly selected from each recording (Fig. 1). Segments were analyzed for noise spectrum level (the sound power in 1-Hz bands of noise, dB *re*: 1 μPa^2/Hz) and total noise-pressure level in the 30–500-Hz band (i.e., the band-pressure level). The spectrum level was determined at intervals of 30 Hz (i.e., 30, 60, 90, 120 Hz, etc.), and also for the test frequencies 70, 100, 200, 400, and 500 Hz used for hearing threshold determinations in a companion study (Lugli *et al.*, 2003). The band-pressure level was calculated as the logarithmic root-mean-square pressure (SPL *re*: 1 μPa). The noise spectrum level and the band-pressure level were computed using the power spectrum and rms functions of AVISOFT, respectively. Decibel values of both noise parameters referenced to 1 V were converted into absolute measurements using the appropriate calibration factors for all components of the measuring system (i.e., hydrophone sensitivity, gain of the DAT recorder, gain of the sound card of the PC). Data from the three noise segments were used to compute the noise spectrum level curve as follows:

$$\mathrm{SL}f = 20 \times \log[\Sigma_i(\exp 10(\mathrm{SL}f, i/20))/3], \quad \mathrm{dB}$$

where SLf is the noise spectrum level at frequency "f", and SL$f,i$ is the noise spectrum level at frequency "f" of the ith noise segment ($i=1-3$). Similarly, the band-pressure level (BL) of the ambient noise was computed using the above formula, with SLf replaced by BL, and SL$f,i$ replaced by BLi, i.e., the band-pressure level of the ith noise segment. The noise spectrum level measurements from each location were used to compute the average spectrum of the stream ambient noise at each frequency ($n=21$) from 30–500 Hz. The mean spectrum level was calculated by averaging the values of SLf (dB) among a given group of locations (quiet and noisy). In addition, the mean spectrum was also calculated only among the group of five locations of the stream with the highest band-pressure levels. The purpose was twofold: to compare the AN close to the heaviest noise sources in the two streams, and to examine the relationship between AN at noisiest places of the stream and the hearing sensitivity of the fish (see Lugli *et al.*, 2003). Besides the mean level, the spectrum level standard deviation (i.e., the standard deviation of SLf) was calculated to estimate the environmental variability of spectrum levels at each frequency. Because only one noise measurement (i.e., only one replicate) was made at each location in the stream, the standard deviation of the mean spectrum levels is an unbiased measure of the environmental variability of the noise level at a given frequency.

B. Noise and sound propagation in Stream Stirone

In addition to single AN measurements at representative sites of the stream, multiple AN measurements were made at two sites of Stream Stirone to study the variability and propagation of AN close to small waterfalls. Variability was determined by recording AN at three locations in a pool below a waterfall (site 1). The three locations were chosen randomly within a small area, about 1 m from the water fall. AN propagation was measured both underwater and above the water surface (site 2) along a transect at 1, 2, 3, and 5 m from a waterfall with depths of 45, 20, 40, and 70 cm, respectively. The transect was positioned to the side of the main channel to minimize the effects of noise generated by rapidly moving water. The air microphone (a miniature condenser microphone with a sensitivity 15 mV/Pa and a frequency response flat in the range 0.01–2 kHz) was manually held about 40 cm above the water surface, with the sensitive element oriented towards the waterfall. The underwater and air recordings were made simultaneously on the left and right channels of the DAT recorder, respectively. As before, three 750-ms segments of the recordings were averaged to quantify spectrum level variability (see figure legends for further details). Spectrum levels (dB *re*: 1 μPa^2/Hz for underwater measurements, or 20 μPa^2/Hz for air measurements) were also computed from a 10-s noise sample (FFT length of 1024 samples, Hamming window and resolution of 3 Hz) at sites 1 and 2 to ascertain whether the differences between locations were real or an artifact of short-term temporal variations.

Propagation of *P. martensii* courtship sounds was also measured in the stream. Sound production was elicited by presenting three territorial males with a ripe, conspecific female inside a small plastic-mesh cage, placed in front of the nest entrance. The caged females elicited courtship sounds, which were monitored by orienting a small, directional Gulton Industries model GLN 9190 hydrophone (sensitivity: -200 dB *re*: 1 μPa, frequency response flat± 1 dB from 10 to 2000 Hz) toward the male. The hydrophone signal was amplified by 40 dB with a Sensor Technology model SA02 preamplifier and recorded onto the DAT. The hydrophone was placed at approximately 5, 20, and 40 cm from the calling male (the exact distance depended upon positioning the hydrophone in relation to the complexity of the bottom), and we recorded multiple sounds for each male at each distance. After the recording, water temperature was measured with a digital thermometer, and the male was netted and measured for total length in millimeters. Sound-pressure level was calculated for the fundamental frequency of the sound as the logarithmic root-mean-square pressure (SPL *re*: 1 μPa) using the power spectrum function of AVISOFT (bandwidth: 10 Hz, Hamming window) and converted to absolute dB (*re*: 1 μPa) using the appropriate calibration factors for all components of the measuring system. Levels of at least five sounds of each male at each distance were averaged to determine transmission loss.

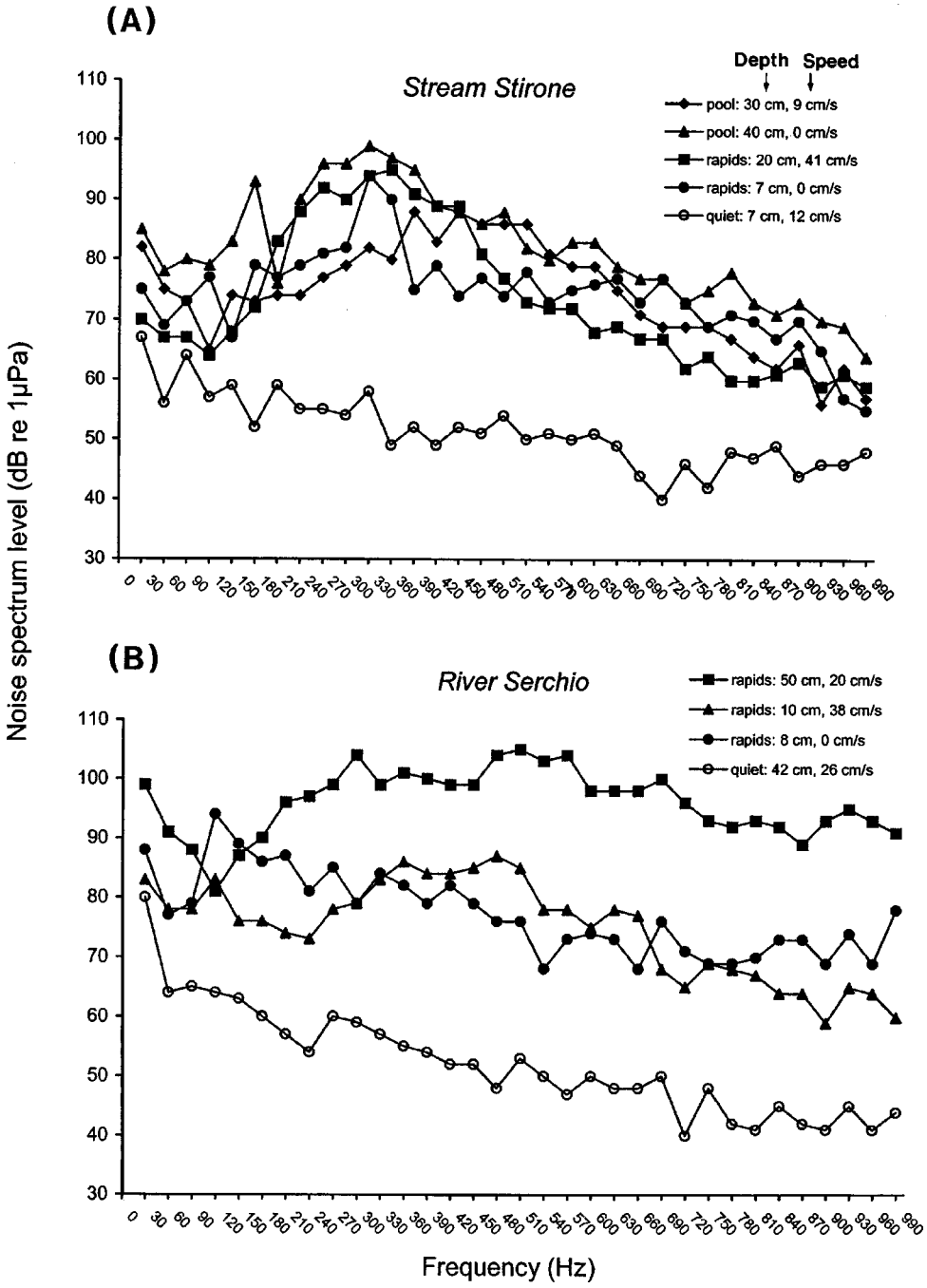


FIG. 2. Ambient noise spectra at four noisy locations and one quiet location from Stream Stirone (A), and at three noisy and one quiet location from River Serchio (B), along with the characteristics of the location (type of location: POOL=hydrophone on the bottom of a pool located below a small waterfall, RAPIDS=hydrophone close to small rapids; water depth, cm; water speed, cm/s). Note spectra of Stream Stirone were similar at noisy locations despite differences in type of location, depth, and water speed, whereas those of River Serchio exhibited greater variability in spectrum shape.

III. RESULTS

A. Stream ambient noise and comparison between streams

Ambient noise (AN) spectra in the frequency range 0.03–1 kHz from quiet locations of both streams have similar shapes and spectrum levels that are usually below 70 dB (*re*: 1 μPa) (Fig. 2). Noise levels are high at low frequencies and fall off with increasing frequency (see also Fig. 3). At noisy locations, however, levels increase at all frequencies, and a variety of spectrum shapes is observed, particularly from River Serchio (Fig. 2). The noise spectra from Stream Stirone are similar in shape, despite differences in the noise source (i.e., waterfall vs rapids), water depth, and current [Fig. 2(A)]. These spectra share a region between 60 and 150 Hz with decreased noise levels. Levels increase to a peak between 300–450 Hz and then decrease by about 5 dB per 100 Hz to 1 kHz. Noisy spectra from River Serchio are more variable than those from Stream Stirone. For example, the curve with highest noise levels [squares, Fig. 2(B)] has a minimum at about 120 Hz, increases by 25 dB to about 500 Hz, and then decreases gradually to 1 kHz. Another curve (circles) has a low-frequency minimum at 60 Hz, a peak at 120 Hz, followed by a decrease to about 600 Hz. The greater variability of AN spectrum shape at River Serchio is statistically supported by converting spectrum level values between 30 and 500 Hz from noisy locations into ranks and comparing them with Kendall's concordance test (Siegel and Castellan, 1988). There is a significant concordance in spectrum shape among locations in Stream Stirone ($W=0.696$, chi-square$=79.3$, $P<0.01$, df$=20$) but not in River

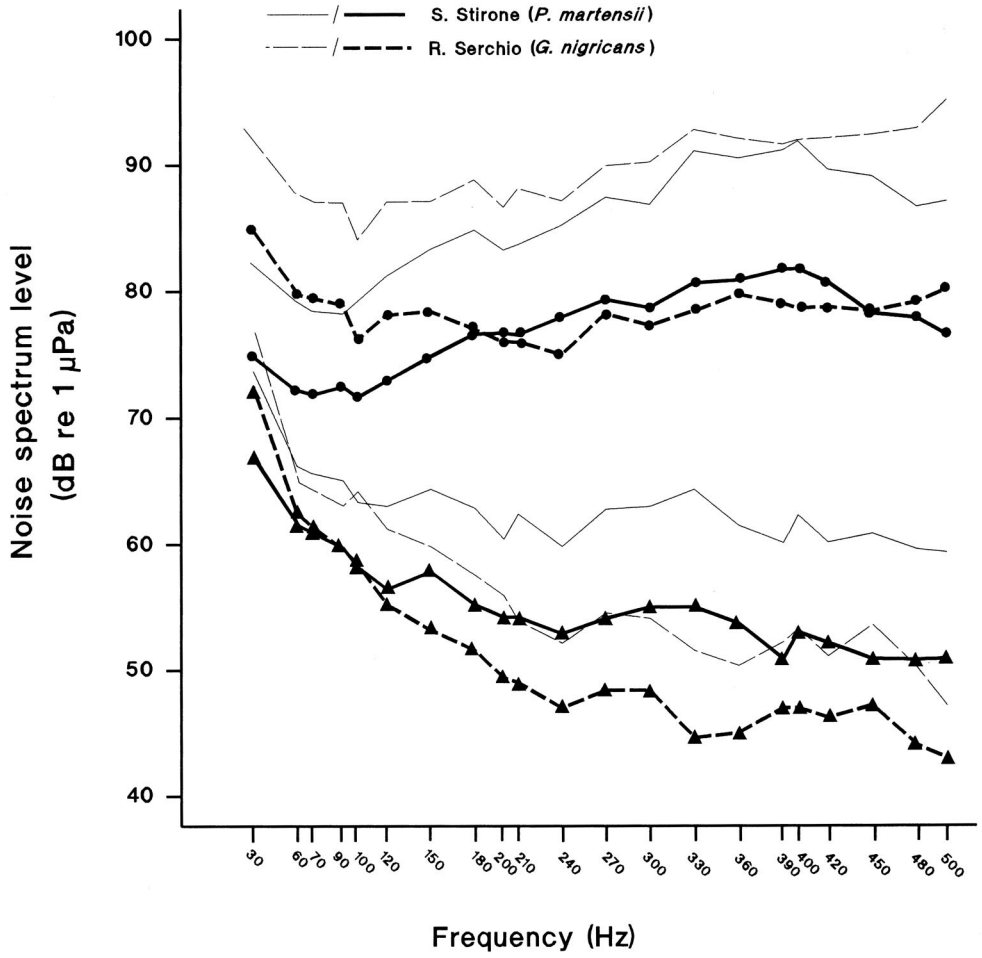


FIG. 3. Mean (thick lines) + 1 s.d. (thin lines) ambient noise spectrum levels at noisy (circles) and quiet (triangles) locations of Stream Stirone and River Serchio. The curve connecting points at 1 s.d. above the mean spectrum level (i.e., the s.d. curve) estimates the environmental variability of the noise levels above the mean level. Notice the flatness of the mean noise spectrum level curve from noisy areas of River Serchio, although a notch at 100 Hz is present in the s.d. curve (see also Fig. 5), and the lower spectrum levels at around 100 Hz (i.e., the quiet window) in the mean spectrum from noisy areas of Stream Stirone.

Serchio ($W = 0.127$, chi-square = 24.0, ns; df = 20). Levels at noisy locations are also different between streams. Although the total noise level from noisy locations in River Serchio ($n = 10$) and Stream Stirone ($n = 13$) does not differ (ns, Mann-Whitney U-test), mean spectrum levels below 100 Hz are 10–15 dB higher in the River Serchio (Fig. 4). The spectrum level differences between the streams decreases from 100 to 150 Hz, and values are similar at higher frequencies (Fig. 3). Note for Stream Stirone, the mean spectrum for noisy locations and particularly for the five noisiest locations (Fig. 4) exhibits a narrow region of lower AN levels around 100 Hz. By contrast, the shape of the mean spectrum for River Serchio is remarkably flat above 30 Hz, regardless of whether it is computed from the ten noisy locations or from the subset of five noisiest locations. This feature of the AN mean spectrum from River Serchio is consistent with the variety of spectrum shapes observed close to sources of AN in this stream. A remarkable feature of the AN at River Serchio is the presence of a ''notch'' at 100 Hz in the s.d. curve of the mean spectrum, particularly from noisiest locations (Fig. 4). The 100-Hz notch may also be a characteristic of individual AN spectra in this stream [see an example in Fig. 2(B)].

B. Short-range propagation and variability of waterfall noise in the Stream Stirone

Noise spectra close to the waterfall at site 1 [Fig. 5(A)] exhibit wide variation in both level and shape despite the short distance between the three selected locations. The frequencies with highest and lowest amplitudes differ among spectra. Note, however, the presence of relatively low noise levels at lower frequencies in two spectra, with minima at 90 and 180 Hz, respectively. Noise spectra determined from 10 s of sampling [Fig. 5(B)] have similar shapes to those in Fig. 5(A), indicating that the spatial differences observed are not due to short-term temporal variations in AN.

The underwater noise spectrum 1 m from the waterfall at site 2 (Fig. 6) has a quiet notch of 82 dB around 100 Hz and maxima around 250 Hz (105 dB) and 420 Hz (104 dB), followed by a gradual but variable decrease in noise level. The noise spectrum from a 10-s sample (Fig. 7, top graph, circles) is similar to the shorter samples (Fig. 6), except for a less deep notch at 100 Hz and a lower peak at 420 Hz. This shape of the AN spectrum is typical for the noisy locations in Stream Stirone (see Fig. 2). Noise spectra from locations 2 or more meters from the waterfall (Figs. 6, 7) have lower spectrum levels at all frequencies and irregular shapes using both procedures for spectrum level calculation (i.e., average value of three 700-ms noise segments or determined from 10 s of sampling). Noise levels at individual frequencies do not vary with distance from the waterfall, except for occasional frequencies (e.g., 30 Hz, 480 Hz, Fig. 6), or narrow frequency ranges (e.g., below 80 Hz, 350–480 Hz, Fig. 7, top graph). Therefore, most of the noise energy from the waterfall attenuates rapidly in these shallow depths (water depth gener-

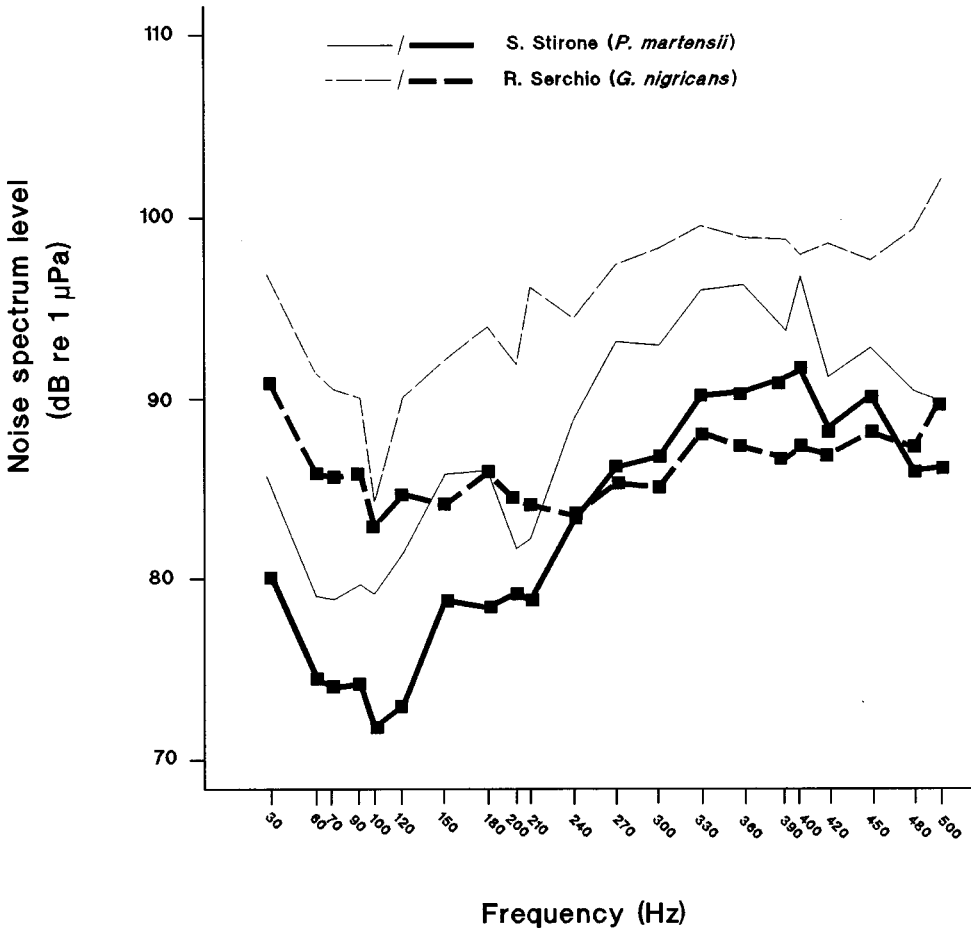


FIG. 4. Mean (thick lines) + 1 s.d. (thin lines) AN spectrum levels from the five noisiest locations of Stream Stirone and River Serchio. Notice the deeper notch at 100 Hz in the s.d. curve, contrasting with the relatively flat mean spectrum level curve, at River Serchio, and the more marked quiet window at Stream Stirone (compare with AN spectra reported in Fig. 3).

ally <50 cm). Note that at frequencies above 540 Hz (Fig. 6, 7), noise levels at the location 5 m from the waterfall tend to exceed those at locations 2 and 3 m from the waterfall. Water height is 70 cm at this location, and 540 Hz is the value of the cutoff frequency for this depth over a rigid bottom (Officier, 1958). Therefore, it is likely that increased noise at higher frequencies is generated by flowing water in the nearby channel. The AN spectrum above the water surface 1 m from the waterfall (Fig. 7, bottom graph) has little or no energy below 200 Hz and maxima around 500 Hz (29 dB), and thereby demonstrates no relationship with the underwater spectrum at the comparable distance (Figs. 6, 7, top graph). In addition, levels of the waterfall noise frequencies exhibit the expected increased attenuation with distance.

C. Propagation of *P. martensii* sounds in Stream Stirone

Sounds were recorded from nest sites in quiet, shallow places (depth< 50 cm) over a stony bottom with low current speeds. The three males were 58-, 72-, and 79-mm TL, and water temperature varied from 19.8 °C to 23.8 °C across recordings. Transmission loss is large in all cases (Fig. 8), with attenuation of 15–20 dB from 5 to 20 cm (all males) and by 30 dB from 8 to 45 cm (one male). No sound is heard with the hydrophone placed 60 cm or more from the nest. Presuming a loss from cylindrical spreading (3-dB/distance doubled), a distance from 5 to 20 cm would account for 6 dB of loss (two doublings). Therefore, 9 to 14 dB of the 15–20-dB loss would be due to absorption within 15 cm, which is equivalent to a loss of 60–90 dB/m.

IV. DISCUSSION OF RESULTS

A. The stream ambient noise

An extensive literature exists on ambient noise in the ocean (Knudsen *et al.*, 1948; Wenz, 1962; Zakarauskas, 1986) and its sources (reviewed in Urik, 1983). Generalized AN spectra of deep water and water over the continental shelf (depth< 200 m) have relatively high noise levels at low frequencies that decrease with increasing frequency (Wenz, 1962). Wide daily and seasonal variations occur at individual locations due to changing weather conditions and types of noise sources. Variability of noise levels is highest in shallow, or very shallow environments, where water depth becomes a critical factor for sound propagation (see below), and wind and wave motion become increasingly important (Urik, 1983). When shipping or biological sources of noise are absent, AN in estuaries is strongly dependent upon wave action, fluctuating widely in relation to prevailing weather conditions. AN levels at low frequencies may be as low as 20–30 dB (pressure spectrum level, dB *re*: 1 μPa) in these very shallow-water environments under calm conditions (Fine and Lenhardt, 1983).

The abundance of literature on AN in the sea contrasts with the paucity of studies in freshwater environments.

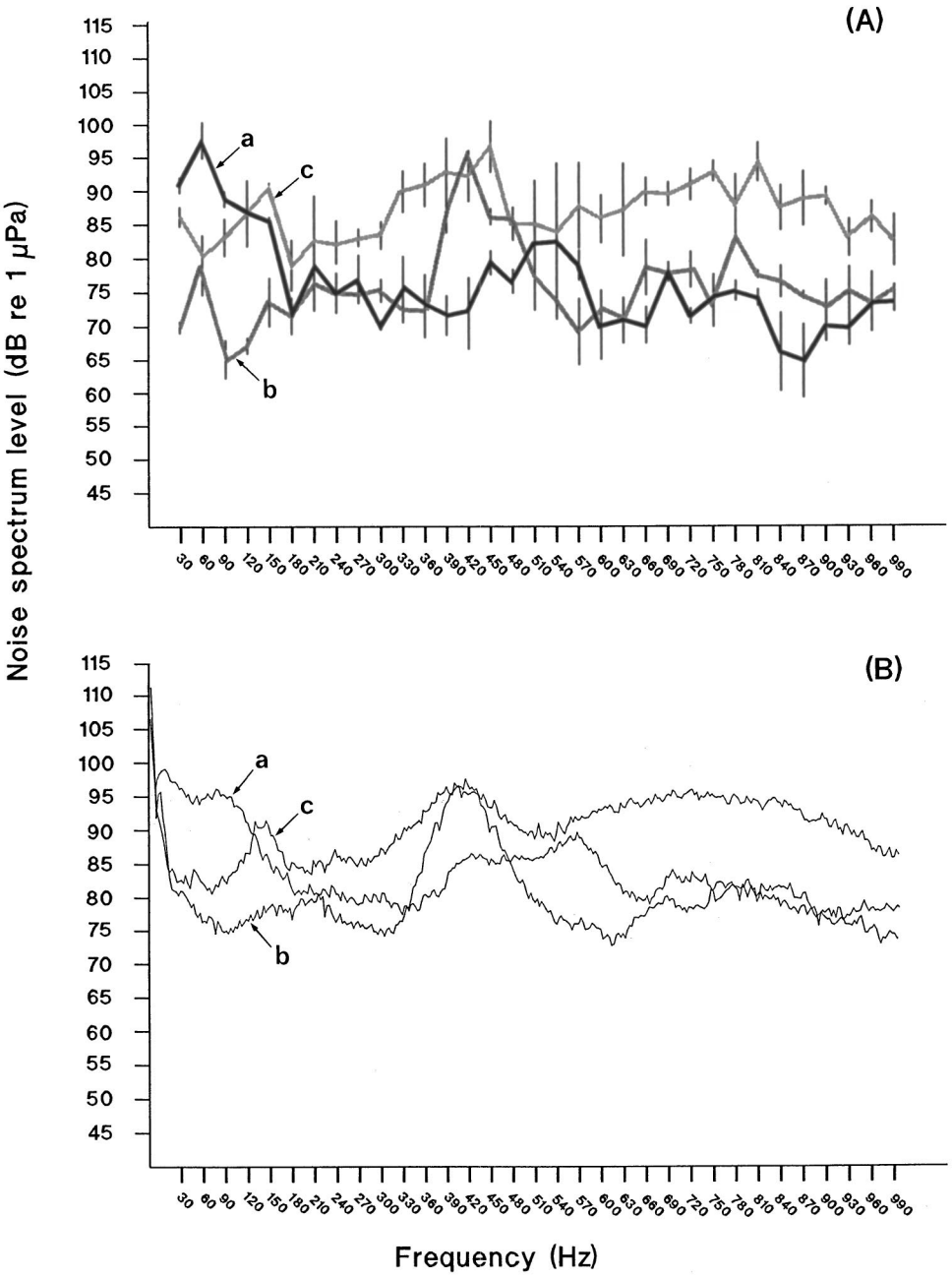


FIG. 5. Variability of AN spectrum in a pool below a small waterfall of the Stream Stirone (site 1). AN was measured at three locations (a, b, and c) chosen randomly within a range of 1.5 m. Water depth was 80, 5, and 15 cm at location a, b, and c, respectively. (A) means (±1 standard error) are SLf values computed by averaging spectrum level measurements from three 700-ms segments as indicated in Sec. II (see also Fig. 1). (B) the AN spectrum at the three locations computed using a 10-s noise segment taken from the same recordings.

Hawkins and Johnstone (1978) found 5–10-dB higher noise levels in the River Dee than in the sea (Loch Torridon) at frequencies of 30–100 Hz, whereas above 150 Hz, and at sea state 3, noise levels were comparable. Highest noise levels measured in the river were around 75 dB *re*: 1 μPa, although Hawkins and Johnstone (1978) predicted even higher levels close to noise sources such as a waterfall.

In the present study, AN spectra were obtained at both quiet and noisy places in two habitats (River Serchio and the small Stream Stirone), which differ in hydrological parameters (stream bed width and average water depth) and bottom characteristics (size and shape of stones on the bottom). However, considering the overlap of AN levels in the streams (total and spectrum noise levels expressed as sound pressure), and the small amount of energy present above 1 kHz, the present measurements are likely representative of rock-lined shallow streams. Regardless of water current, the streams are remarkably quiet in places where the water surface is unbroken (noise spectrum levels from 40 to 60 dB *re*: 1 μPa). Such places are fairly common in the two streams because of the modest slope of the stream bed (Lugli *et al.*, 1992, for Stream Stirone). When the water surface breaks, trapping air underwater, background noise increases significantly at all frequencies, with highest increments in the 200–500-Hz band. Noisiest places in the stream (i.e., total SPL in the frequency band 30–500 Hz: 110–130 dB *re*: 1 μPa) occur near small rapids or in pools below a small waterfall. The spectrum levels at such places may differ up to 35 dB between different frequencies in the 0.03–1-kHz band of a spectrum, and they often exceed levels measured in shallow ocean at higher sea states or during heavy shipping by 10–15 dB. Furthermore, the AN spectrum may vary remarkably in both shape and level among nearby locations ($<$1.5 m) close to a waterfall. Large variations in AN levels between nearby

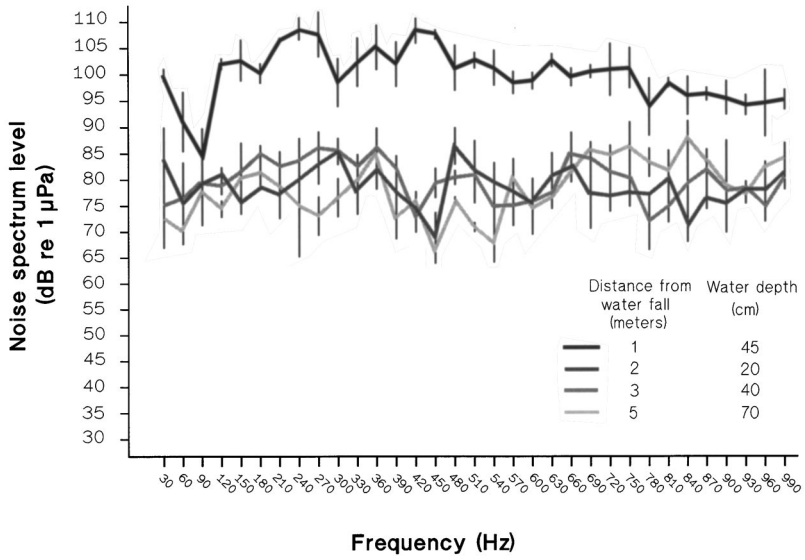


FIG. 6. Variation of AN spectrum with distance from a waterfall in Stream Stirone (site 2). AN was measured at four locations along a transect line downstream from 1–5 m from the waterfall. The bottom of the study site was mainly pebbles and stones and varied from 20–70 cm in depth among the four locations. Means and standard errors of the underwater AN spectrum computed from three 700-ms segments as indicated in Sec. II.

locations are expected because lower frequencies do not propagate at shallow depths and decay exponentially with distance from the source, i.e., the frequency cutoff phenomenon (Officier, 1958; see below).

Our measurements of AN propagation in Stream Stirone indicate that most energy of the low-frequency noise generated underwater by a waterfall is lost within only 2 m of the fall, whereas the airborne noise from the same waterfall propagates above the water surface several meters away from the source. Furthermore, regardless of the distance (i.e., 1, 2, 3, or 5 m) from the waterfall, there is little resemblance between the waterborn and airborne noise spectra at the same location, indicating that little of the acoustic energy from the waterfall noise in air is transmitted to the water. Low transmission is expected because of the difference in acoustic impedance between air and water and because of the large angles (near 90°, in this case) between the noise source and the receiver, i.e., Snell's law (Urik, 1983).

A previously undescribed feature of stream AN is a quiet window around 100 Hz in many noisy locations (particularly in Stream Stirone). The window is about 130 Hz wide in Stream Stirone (Figs. 3, 4), and it is more sharply tuned at 100 Hz, i.e., resembling a notch, in River Serchio (Fig. 4).

The 100-Hz ''notch'' in the AN spectrum occurs between two sources of water noise, which we attribute to water turbulence at the low end, and waterfalls and rapids, the equivalent of wave action in the ocean at the high end (Urik, 1983). Water turbulence is the likely source of AN (i.e., turbulence noise) at frequencies below 100 Hz since underwater currents in the sea generate turbulence noise in the 1–100-Hz band (Wenz, 1962; Urik, 1983). In the stream, propagation of low frequencies is further constrained by the low water depths (Urik, 1983). Therefore, low-frequency noise originating from water turbulence is likely to be a significant component of stream AN at places near or inside a turbulence zone. Two findings of the present study support this

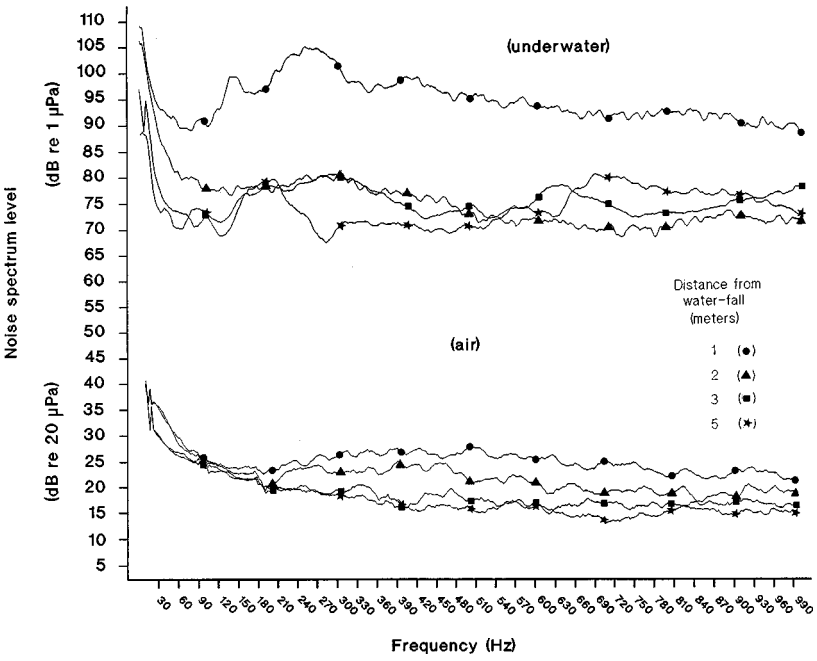


FIG. 7. Top graph: the underwater spectrum at the four locations of Fig. 6 computed using a 10-s noise segment taken from the same recordings. Bottom graph: the spectrum of the waterfall noise propagating above the water surface computed using a 10-s noise segment taken from recordings with the air microphone at the four locations.

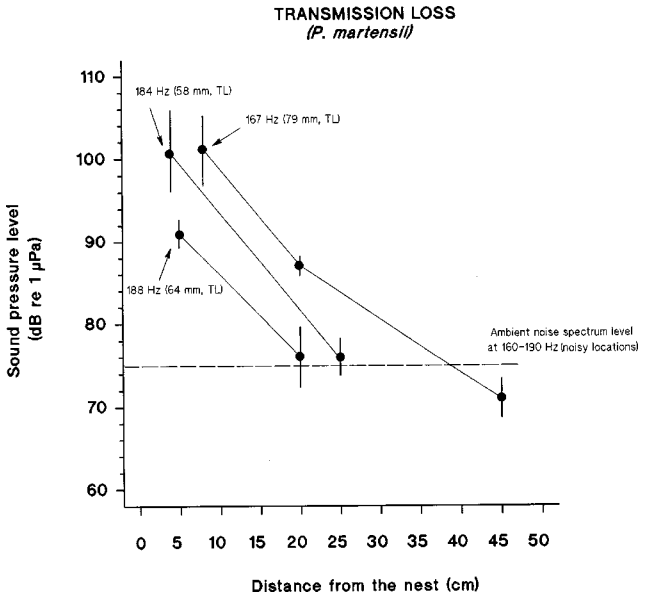


FIG. 8. Transmission loss with distance (cm) for the fundamental frequency of sounds emitted by three male *P. martensii* courting a conspecific, ripe female in the field (Stream Stirone). Sounds were emitted by the male at the nest entrance. Mean pressure levels (dB *re*: 1 μPa) of the fundamental frequency (at around 180 Hz) were computed from five sounds at each distance. Differences in fundamental frequency between males are explained by water temperature (Torricelli *et al.*, 1990). In addition to the sound levels, the mean ambient noise spectrum level at 180 Hz from the noisy locations of Stream Stirone (see Fig. 3) is reported for comparison.

hypothesis. First, measurements of propagation in Stream Stirone indicated that AN levels at frequencies below 100 Hz decrease with distance from the waterfall (see Figs. 6, 7), a result consistent with decreasing water turbulence at greater distances. Second, River Serchio has a higher water discharge and, presumably, higher pressure changes associated with water turbulence than Stream Stirone, which is consistent with higher low-frequency noise levels measured in River Serchio (see Fig. 3).

Small waterfalls and rapids are likely sources of stream AN at frequencies above 100 Hz. Prosperetti (1985) and Carey (1985) suggested that entrainment of clouds of air bubbles below the surface during wave action may be a significant source of energy of AN below 1 kHz in the sea. Laboratory investigations indicated main energy of collective bubble cloud oscillations resting in the 200–600-Hz band (Yoon *et al.*, 1991). The stream environment is characterized by abundant water splashes and formation of air bubbles underwater, with most of the energy close to waterfalls and rapids falling in the 200–500-Hz band. Presumably, water splashes and air bubbles are important interrelated sources of low-frequency ambient noise in the stream, as are wind-dependent bubble and spray noise in the ocean (Franz, 1959; Wenz, 1962; Prosperetti, 1988).

The combined effects of turbulence noise and bubble noise, as inferred from studies on sources of oceanic ambient noise and observations of the present study, would leave a narrow region (a ''notch'') of relatively low noise levels around 100 Hz, a feature common to the mean spectrum from noisy locations of both streams.

B. Sound propagation in the stream

Acoustics of very shallow water (estuaries, rivers, and ponds) is more complex than of deeper waters because of the proximity of the surface and the bottom (Rogers and Cox, 1988). Additionally, water depth strongly constrains low-frequency communication in shallow water because no frequency corresponding to a wavelength longer than approximately four times the water depth will propagate underwater, i.e., the frequency cutoff phenomenon over a rigid bottom (Officier, 1958; Rogers and Cox, 1988). For example, frequencies below 750 Hz will not propagate in water with a depth of 50 cm or less, and *P. martensii* and *G. nigricans* emit sounds with main frequencies in the 80–200-Hz band (Lugli *et al.*, 1995, 1996b, and Lugli *et al.*, 2003). These frequencies are well below the cutoff frequencies of the stream (0.8–7 kHz for water depths from 5 to 50 cm) over a rigid bottom (Officer, 1958). Our field measurements of courtship sound transmission in *P. martensii* indicate an attenuation of 15–20 dB over 20 cm at depths $<$50 cm. Due to the low amplitude of these sounds (90–120 dB at 5–10 cm, Lugli *et al.*, 1995; Lugli *et al.*, 2003), calls are lost in noise 50–60 cm from the source, even under quiet conditions. Attenuation is far higher than previously measured for fish sounds, and is similar to values of aquatic insects singing at frequencies $>$2 kHz in 21-cm-deep freshwater (Aiken, 1982).

The effects of the high transmission loss of the sound in the stream, the low sound amplitude and auditory sensitivity of *P. martensii* and *G. nigricans* (see Lugli *et al.*, 2003) combine to restrict acoustic communication in these species to only a few decimeters from the calling male. Playback sounds broadcast to *P. martensii* in laboratory tanks may attract individuals (aggressively aroused males or ripe females) as far as 40 cm away from the speaker (Lugli *et al.*, 1996a; Lugli, 1997), and they are probably not heard at distances greater than 50 cm (Lugli, 1997, personal observation). However, since the average distance between neighboring male *P. martensii* in the stream is also around 60 cm (Lugli *et al.*, 1992), and females are interspersed among male territories, there is probably no need for long-distance acoustic communication in this species.

Short-range communication in territorial shallow-water fishes (freshwater, estuaries, coastal waters) has been reported by other authors (Tavolga, 1958; Gerald, 1971; Fine, 1981; Fine and Lenhardt, 1983; Mann and Lobel, 1997) although propagation occurred over several meters. Attenuation was not as extreme as in the goby streams because some of the sound energy was above the cutoff frequencies in these environments. Furthermore, softer sand and mud bottoms, unlike the stones in the current study, appear to permit some degree of propagation, i.e., they don't appear to function as a rigid boundary.

Studies on acoustic communication in terrestrial environments have shown that animals have coped with environmental constraints on sound transmission by evolving sound features that maximize long-range communication (Waser and Waser, 1977; Brenowitz, 1982; Wiley and Richards, 1982; Ryan and Brenowitz, 1985; Klump, 1996). The importance of the environment on acoustic communication and

sound features among fishes is far less clear (Fine and Lenhardt, 1983; Roger and Cox, 1988; Forrest *et al.*, 1993; Bradbury and Veherencamp, 1998). This study establishes that depth and AN level are the major constraints on acoustic communication range in the stream gobies. A companion paper (Lugli *et al.*, 2003) further demonstrates that these factors appear to represent strong selective forces on the evolution of sound frequencies produced by these fishes and their hearing sensitivity.

ACKNOWLEDGMENTS

This study was supported by grants from MURST (Italian Ministry of Scientific Research and University) and from the Office of Naval Research (USA). We are grateful to Romano Romani for help in data collection and technical assistance, and to Tim Cameron and Raimund Specht for comments on an earlier version of the manuscript.

Aiken, R. B. (**1982**). ''Shallow-water propagation of frequencies in aquatic insect sounds,'' Can. J. Zool. **60**, 3459–3461.

Bradbury, J. W., and Veherencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer, Sunderlands).

Brenowitz, E. A. (**1982**). ''Long-range communication of species identity by song in the red-winged blackbird,'' Behav. Ecol. Sociobiol. **10**, 29–38.

Carey, W. M. (**1985**). ''Low-frequency ocean surface ambient noise,'' J. Acoust. Soc. Am. Suppl. 1 **78**, S1–S2.

Fay, R. R., and Popper, A. N. (**1999**). Comparative hearing: fish and amphibians (Springer, New York).

Fine, M. L. (**1981**). ''Mismatch between sound production and hearing in the oyster toadfish,'' in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 257–263.

Fine, M. L., and Lenhardt, M. L. (**1983**). ''Shallow-water propagation of the toadfish mating call,'' Comp. Biochem. Physiol. A **76**, 225–231.

Fine, M. L., Winn, H. E., and Olla, B. L. (**1977**). ''Communication in fishes,'' in *How Animals Communicate*, edited by T. A. Sebeok (Indiana University Press, Bloomington), pp. 472–518.

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.

Franz, G. J. (**1959**). ''Splashes as sources of sound in liquids,'' J. Acoust. Soc. Am. **31**, 1080–1096.

Gandolfi, G., and Tongiorgi, P. (**1974**). ''Taxonomic position, distribution and biology of the gobies present in the Italian fresh-water, *Padogobius martensii* (Günther) and *Gobius nigricans* Canestrini (Osteichtyes, Gobiidae),'' Ann. Mus. Civ. St. Nat. Genova **80**, 92–118.

Gerald, J. W. (**1971**). ''Sound production during courtship in six species of sunfish (*Centrarchidae*),'' Evolution (Lawrence, Kans.) **25**, 75–87.

Hawkins, A. D. (**1981**). ''The hearing abilities of fish,'' in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 109–137.

Hawkins, A. D., and Johnstone, A. D. F. (**1978**). ''The hearing of the Atlantic Salmon, *Salmo salar*,'' J. Fish Biol. **13**, 655–673.

Hawkins, A. D., and Myrberg, A. A., Jr. (**1983**). ''Hearing and sound communication under water,'' in *Bioacoustics: A Comparative Approach*, edited by B. Lewis (Academic, London), pp. 347–405.

Klump, G. M. (**1996**). ''Bird communication in the noisy word,'' in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Comstock, Cornell University Press, Ithaca and London), pp. 321–338.

Knudsen, V. O., Alford, R. S., and Emling, J. W. (**1948**). ''Underwater ambient noise,'' J. Mar. Res. **3**, 410–429.

Lugli, M., (**1997**). ''Response of male goby, *Padogobius martensil*, to agressive sound playback follwing pre-experimental visual stimulation''. Behav. **134**, 1175-1188.

Lugli, M., Bobbio, L., Torricelli, P., and Gandolfi, G. (**1992**). ''Breeding ecology and male spawning success in two hill-stream populations of the freshwater goby, *Padogobius martensii*,'' Environ. Biol. Fishes **35**, 37–48.

Lugli, M., Pavan, G., Torricelli, P., and Bobbio, L. (**1995**). ''Spawning vocalizations in male freshwater gobies (Pisces, Gobiidae),'' Environ. Biol. Fishes **43**, 219–231.

Lugli, M., Pavan, G., and Torricelli, P. (**1996a**). ''The importance of breeding vocalizations for mate attraction in a freshwater goby with composite sound repertoire,'' Ethol. Ecol. Evol. **8**, 343–351.

Lugli, M., Torricelli, P., Pavan, G., and Miller, P. J. (**1996b**). ''Breeding sounds of male *Padogobius nigricans* (Teleostei: Gobiidae) with suggestions for further evolutionary study of vocal behavior in gobioid fishes,'' J. Fish Biol. **49**, 648–657.

Lugli, M., Torricelli, P., Pavan, G., and Mainardi, D. (**1997**). ''Sound production during courtship and spawning in freshwater gobies (Pisces, Gobiidae),'' Mar. Freshw. Behav. Physiol. **29**, 109–126.

Lugli, M., Yan, H. Y., and Fine, M. L. (**2003**). ''Acoustic communication in two freshwater gobies: The relationship between ambient noise, hearing thresholds, and sound spectrum.'' J. Comp. Physiol., A (in press).

Mann, D. A., and Lobel, P. S. (**1997**). ''Propagation of damselfish (*Pomacentridae*) courtship sounds,'' J. Acoust. Soc. Am. **101**, 3783–3791.

Myrberg, Jr., A. A. (**1981**). ''Social communication and interception in fishes,'' in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 395–425.

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.

Officier, C. B. (**1958**). *Introduction to the Theory of Sound Transmission* (McGraw-Hill, New York).

Prosperetti, A. (**1985**). ''Bubble-related ambient noise in the ocean,'' J. Acoust. Soc. Am. Suppl. 1 **78**, S2.

Prosperetti, A. (**1988**). ''Bubble-related ambient noise in the ocean,'' J. Acoust. Soc. Am. **84**, 1042–1054.

Rogers, P. H., and Cox, M. (**1988**). ''Underwater sounds 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, New York), pp. 131–149.

Ryan, M. J., and Brenowitz, E. A. (**1985**). ''The role of body size, phylogeny, and ambient noise in the evolution of bird song,'' Am. Nat. **126**, 87–100.

Siegel, S., and Castellan, N. J., Jr. (**1988**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).

Sokal, R. R., and Rohlf, F. J. (**1981**). *Biometry* (Freeman, New York).

Tavolga, W. N. (**1958**). ''The significance of underwater sounds produced by males of the gobioid fish *Bathygobius soporator*,'' Physiol. Zool. **31**, 259–271.

Torricelli, P., Lugli, M., and Pavan, G. (**1990**). ''Analysis of sounds produced by male *Padogobius martensi* (Pisces, Gobiidae) and factors affecting their structural properties,'' Bioacoustics **2**, 261–275.

Urik, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula, Los Altos, CA).

Waser, P. M., and Waser, M. S. (**1977**). ''Experimental studies of primate vocalization: Specializations for long distance propagation,'' Z. Tierpsychol **43**, 239–263.

Wenz, G. M. (**1962**). ''Acoustic ambient noise in the ocean: Spectra and sources,'' J. Acoust. Soc. Am. **34**, 1936–1956.

Wiley, R. H., and Richards, D. G. (**1982**). ''Adaptations for acoustic communication in birds: Sound transmission and signal detection,'' in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Academic Press), Vol. 1, pp. 131–278.

Yoon, S. W., Crum, L. A., Prosperetti, A., and Lu, N. Q. (**1991**). ''An investigation of the collective oscillations of a bubble cloud,'' J. Acoust. Soc. Am. **89** (2), 700–706.

Zakarauskas, P. (**1986**). ''Ambient noise in shallow water: A literature review,'' Can. Acoust. **14**, 3–17.

A diminished rate of "physiological decay" at noise offset contributes to age-related changes in temporal acuity in the CBA mouse model of presbycusis

James R. Ison[a] and Paul Allen
Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, New York 14627

(Received 18 October 2002; accepted for publication 7 April 2003)

The persistence of afferent activity at stimulus offset was examined in mice as a function of age (experiment 1) and noise level (experiment 2), using a procedure devised by von Bekesy to study the physiological decay of afferent activity. Noise offset was presented from 1 to 10 ms (the temporal gap) prior to an acoustic startle stimulus, with an abrupt edge or a linear ramp having the duration of the gap. Noise offset inhibited the startle reflex, this effect declining with age. For young (2–3 months old) and ''young-old'' mice (18–19 months old), the abrupt offset was always more inhibitory than the ramp, even for the one millisecond gap. In ''old-old'' mice (24–26 months old) abrupt and ramped offsets were not different until the gap exceeded 4 ms. The behavioral difference between abrupt and ramped decay times did not depend on noise level in young mice (4–5 months old), though overall inhibition increased with level. These data indicate that temporal acuity as measured by this gap detection method is limited in very old mice by their reduced ability to follow the abrupt change in noise level at the beginning of the gap. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1577553]

PACS numbers: 43.80.Lb, 43.66.Gf, 43.66.Mk [WPS]

I. INTRODUCTION

Many years ago von Bekesy (1933/1960) measured the slowest rate of tone decay that a human listener was unable to distinguish from a tone with an abrupt offset, in order to track the rate of ''physiological decay'' at stimulus offset in human listeners. He appreciated that neural activity does not immediately cease at stimulus offset, and attributed the sensory equivalence of the two tones to their having the same internal rate of decay of afferent activity: for the tone with the abrupt offset this was the intrinsic decay of neural excitation, while for the tone with the ramped offset the more gradual decrement in acoustic input matched the rate of the internal decay process. A later experiment by Miller (1948) extended this procedure more appropriately to noise offsets, which do not yield spectral differences for abrupt and ramped transitions, and found that the slowest linear decay time that could not be distinguished from an abrupt shift was about 35 ms for his human listener, this value being invariant across noise levels. Plomp (1964) followed up these earlier experiments in his study of ''the decay of auditory sensation,'' but developed the gap detection method for this research in preference to the comparison of abrupt versus ramped offsets previously employed by von Bekesy (1933/1960) and Miller (1948). Gap detection experiments are now the most commonly used method for studying temporal acuity, and yield thresholds of about 2 to 3 ms under optimal conditions for both human listeners and for laboratory animals.

In a previous study of the effects of age on temporal acuity in humans and in mice, Barsz *et al.* (2002) reported that old CBA mice and old human listeners with near-normal audiograms are similar in their being less sensitive to gaps than are the young, and provided single-unit electrophysiological data from the mouse inferior colliculus suggesting that this loss of sensitivity can be attributed to an age-related deficit in the strength of the onset response to noise at the end of the gap. But as described first by Plomp (1964), gap detection depends on two partially independent processes, namely, first, the rate of decay of excitation following noise offset at the initial edge of the gap, and, second, the strength of the excitatory response to noise onset at its end. The present experiment was designed to test the hypothesis that the decline in temporal acuity in the old subject may result at least in part from an age-related increase in the persistence of neural activity during the gap.

In the present research we returned to the procedure initially proposed by von Bekesy in combination with an adaptation of the method of reflex modification audiometry (Young and Fechter, 1983) to investigate temporal acuity in the mouse. An abrupt noise offset provides a powerful graded inhibition of the acoustic startle reflex (ASR) with an increase in the duration of the gap that intervenes between noise offset and the startle stimulus in mice (Ison *et al.*, 1998). A ramped noise offset that intrudes into a gap reduces gap detection, in mice (Ison *et al.*, 2002) and in humans (Allen *et al.*, 2002). These data indicate that at long decay times the abrupt decay should produce more inhibition than the gradual decay, but, further, in line with the hypothesis of von Bekesy, at some relatively short decay rates that these offset conditions may be indistinguishable in their effects. The question then is how the temporal threshold for detecting this difference between the abrupt and ramped offsets might vary with age.

[a]Electronic mail: ison@bcs.rochester.edu

 0001-4966/2003/114(1)/522/7/$19.00

Three age groups were tested, young adult, old, and very old mice. If the rate of decay of afferent stimulation varies with age, then the two noise offset functions should diverge at different gap durations in the three groups. A second experiment investigated the effect of differences in noise level on these functions. At two years of age old CBA mice have a 20- to 30-dB loss of sensitivity in their hearing across the spectrum (Li and Borg, 1991), and an effect of age on the rate of decay might be an indirect consequence of hearing loss rather than its having a direct link to age. Although neither Plomp (1964) nor Miller (1948) had found a difference in decay rate across level, the experimental details of their work were very different from those used here. For this reason we tested a single group of young mice in three different background noise levels, to determine whether the temporal functions for the development of startle inhibition follow abrupt versus ramped noise offsets would vary with noise level.

II. METHOD

A. Subjects

The subjects were 102 CBA/CaJ mice (54 male, 48 female) born and raised in the vivarium at the University of Rochester, derived from breeding stock initially acquired from the Jackson Laboratory (Bar Harbor, ME). In experiment 1 the mice were 2–3 months ($n=26$), 18–19 months ($n=31$), or 24–26 months of age ($n=25$). In experiment 2 the mice were 5–6 months of age ($n=20$). The mice were maintained in group cages under conditions of constant temperature and humidity, with a 12–12 h light dark cycle (lights on at 6 a.m.). Testing was conducted usually between the hours of 9 a.m. to 6 p.m.

B. Apparatus

The details of the apparatus and general procedures have been previously described (Ison *et al.*, 2002), including (in Fig. 1 of that report) a depiction of the startle response on baseline control trials and on trials in which the startle stimulus is preceded by a gap in a background noise. Briefly, the mouse was confined for testing in a wire mesh oval-shaped cage mounted on a suspended acrylic platform directly over an accelerometer, this assembly placed in an anechoic chamber. The accelerometer was sensitive to the vertical force exerted by the startle reflex, and its output was amplified and then integrated over a 100-ms interval beginning with the startle stimulus. The startle stimulus was a noise burst presented at 115 dB SPL, and was 20 ms in duration with near instantaneous rise and fall times (<0.2 ms). To human listeners this stimulus was about equivalent to a finger-snap into the palm from a distance of about 5 cm. In experiment 1 the subjects were tested in the presence of a 70 dB SPL noise background, and in experiment 2 the subjects were tested in noise backgrounds of 30, 50, and 70 dB SPL on separate test days. The startle stimulus and the background noise were presented through separate high-frequency tweeters. The background noise was controlled by an electronic switch that gated the noise off abruptly or gradually just prior to the presentation of the startle stimulus.

C. Procedure

The mice received trials on which the startle stimulus was delivered alone in the noise background with no preceding stimulus event and trails in which the startle stimulus was preceded by noise offset, the beginning of the offset leading to onset of the startle stimulus by 1, 2, 3, 4, 5, or 10 ms. On half of these prestimulus trials the noise ended abruptly (nominally 0 ms, actually with a decay time <0.2 ms) and on half of these trials the noise level was ramped off and the decay time (DT) of the linear ramp equaled the offset lead time (LT) for that trial. Thus all of the startle stimuli on these prestimulus trials were presented in quiet, in one set of trials after silent periods of 1, 2, 3, 4, 5, 10 ms (DT=0) and in the other set of trials immediately at the end of a ramped offset that had a duration of 1, 2, 3, 4, 5, or 10 ms (DT=LT). There were 15 stimulus conditions in total: 2 startle stimulus-alone conditions that provided the baseline control level; 12 prestimulus conditions with 6 lead times given for each of the two types of decay, ramped versus abrupt; and a single condition in which no stimuli were given but activity was measured for 100 ms at the time that a startle stimulus would normally have been given. These conditions were given in random order within blocks of 15 trials, with an average of 20 s between trials (range 15 to 25 s). A total of 11 blocks of trials were given in a test day, and the first block was discarded in order to minimize the effect of novelty that would affect performance on the first few trials. Mice were tested one at a time, and the total test time was about 1 h in duration. In experiment 1 two test days were scheduled, at least 1 day apart, with the background noise set at 70 dB for each test. In experiment 2 three test days were given, on average 4 days apart, the background noise levels variously set at 30, 50, or 70 dB, in counterbalanced order across days.

D. Data analysis

Two indices of response strength were used for the primary statistical analyses. The first was the mean integrated response amplitude on no-prestimulus control trials, in units linearly related to the force of the response and measured in voltage units. This measure provided information about the effects of age on the vigor of the startle reaction. It also provided the critical startle baseline value against which reflex modification by the prestimulus is assessed. The second was the relative response amplitude on prestimulus trials, this based on the ratio formed by dividing the mean response amplitude for each mouse for each stimulus condition by the means of its control response amplitude. Both response measures were subjected to the analysis of variance (ANOVA). In the first experiment age was a between-subjects variable, and lead time (1, 2, 3, 4, 5, and 10 ms) and decay time (abrupt or ramped) were within-subject variables. In the second experiment lead time, decay time, and noise level (30, 50, and 70 dB) were all within-subject variables. Degrees of freedom for repeated measured were adjusted for nonhomogeneity of between-cell correlations by the Hunyh-Feldt procedure. The major significant findings from the ANOVA are shown in Table I.

TABLE I. Major ANOVA results for experiment 1.

Source	ASR baseline	Relative response
Age	$F(2/76)=57.70$[a]	$F(2/76)=7.61$[a]
Lead Time		$F(5/380)=193.54$[a]
Delay Time		$F(1/76)=143.24$[a]
A×LT		$F(10/380)=4.65$[a]
A×DT		$F(1/76)=23.50$[a]
LT×DT		$F(5/380)=7.26$[a]

[a]$p<0.01$.

III. RESULTS

A. Startle response amplitude

1. Experiment 1

The amplitude of the baseline startle reflex was systematically reduced with age, the mean values in voltage units (±SEM) being 4795 (311), 3079 (216), and 954 (86), at 2–3, 18–19, and 24–26 months of age, respectively. This difference was significant (see Table I). Males responded more vigorously than females, this difference being of marginal significance, $F(1/76)=3.70$, $p=0.058$. There were no significant differences apparent between the groups in background activity on trials that did not have a startle stimulus, all $F<1.0$. There was thus a greater difference between the mean startle response and the activity level in younger mice, with mean differences (±SEM) of 4537 (317), 2797 (220), and 666 (84) from the youngest to the oldest group, respectively. However, the difference in the oldest group sufficiently large in SEM units (with $t=7.89$, $df=42$) that the activity level would not create an artificial floor on the potential strength of reflex inhibition.

2. Experiment 2

The vigor of the control startle response was greatest in the presence of the 50 dB noise background for these mice, with a mean of 3519 (276), compared to means of 2995 (251) in the 70-dB background and 2397 (272) in the 30-dB background. These difference were significant (see Table II), and the ANOVA also provided a significant quadratic trend ($p=0.0009$) that reflected the effect of the nonmonotonic ASR function across level. Subsequent *t*-tests found that the reduced response levels at both 70 and 30 dB compared to 50 dB were significant, $p<0.05$. There was no difference between the amplitudes of the startle reflex in males and females, $F<1$.

TABLE II. Major ANOVA results for experiment 2.

Source	ASR baseline	Relative response
Level	$F(2/36)=9.36$[a]	$F(2/36)=23.46$[a]
Lead Time		$F(5/90)=119.48$[a]
Decay Time		$F(1/18)=90.49$[a]
L×LT		$F(10/180)=2.93$[a]
L×DT		$F(2/36)=0.06$

[a]$p<0.01$.

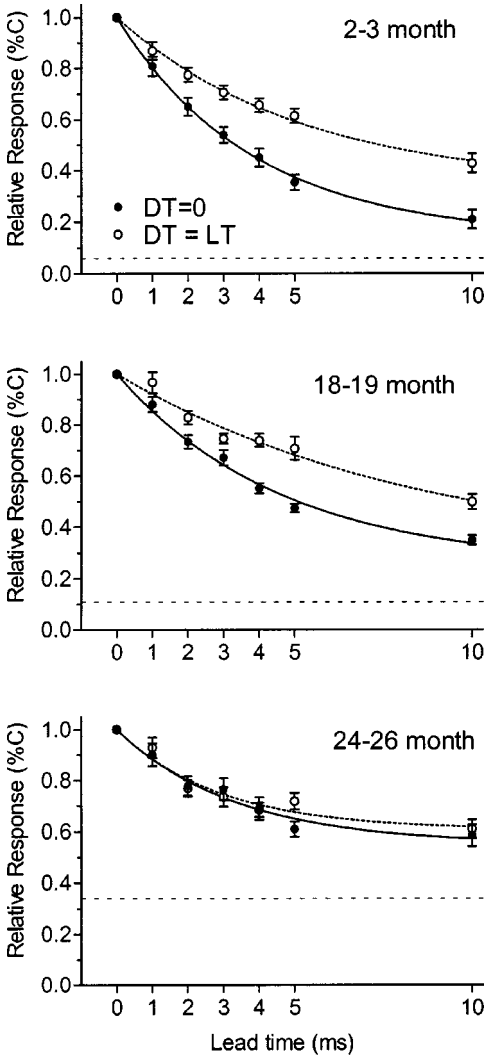


FIG. 1. Mean relative amplitude (±SEM) for the startle reflex when the eliciting stimulus was presented alone (the control level at LT=0) or when preceded by noise offset at the depicted lead times, this being the duration of the brief gap between noise offset and the onset of the startle stimulus. In one condition the decay time of the offset was abrupt (DT=0) and in the other condition the linear ramped decay time was the same duration as the lead time (DT=LT). The lines describe the fit to the data provided by an exponential decay function with two parameters, the time constant and the asymptotic plateau. The horizontal dashed line is the background activity level in the absence of any stimulus. The performance of mice at three ages is shown.

B. Startle response modification by noise offset

1. Experiment 1

Figure 1 depicts the mean (±SEM) relative response values as a function of lead time (LT) for the three age groups, when the decay time (DT) of noise offset was abrupt or was ramped down over the duration of the gap. The lines, solid for DT=0 and broken for DT=LT, are the product of a nonlinear regression analysis that passed through 1.0 at LT=0 and provided a best least-squares fit to the data points with an exponential decay function. The decline in response strength in the DT=0 condition approximated the functions obtained for CBA mice of similar ages by Ison *et al.* (1998), showing systematic differences in the asymptotic plateau with increasing age (0.14, 0.26, and 0.55 from youngest to oldest), but not in the time constant (3.8, 4.5, and 3.3 ms).

The onset of inhibition was very rapid in all three age groups, so that the levels of the startle response were significantly reduced compared to the control value when the lead time of the abrupt noise offset was just 1 ms: by 19% for the 2–3 month-old mice, $t(25)=5.21$, $p<0.0001$; by 12% for 18-month-old mice, $t(30)=4.04$, $p=0.0003$; and by 10% for the 24–26-month-old mice, $t(24)=2.21$, $p=0.037$. The major findings of the overall ANOVA of the relative response data given in Fig. 1 are provided in Table I, showing significant main effects of age, lead time, and decay time, and interactions of age with LT and with DT, and an interaction of LT and DT. The near asymptotic level of inhibition obtained at 10 ms when DT=0 was reduced with increasing age ($p<0.0001$). Although the mean activity score of the oldest mice was much higher relative to the level of the control ASR than it was in the younger mice (at 34%, compared to 6% and 11%) it was still much lower than the near asymptotic relative response at DT=0 and LT=10, $t(24)=5.54$, $p<0.0001$. Separate ANOVA of the three age groups provided significant main effects for lead time, decay time, and a LT×DT interaction for the 2–3 and 18–19-month-old mice (all $p<0.0001$), but for the 24–26-month-old mice there was a significant main effect for lead time ($p>0.0001$) but no significant effects attributable to the difference between ramped abrupt offsets ($p>0.25$). The pattern of effects was the same in male and female mice. As is apparent in Fig. 1, the effect of ramped decays in the noise compared to the abrupt offset was to reduce the overall level of inhibition in the 2–3- and 18–19-month-old mice, but not in the oldest group of mice, at 24–26 months of age. The difference between the two types of noise offset decay increased with lead time, but in an ANOVA of the 2–3- and 18–19-month-old mice the difference between DT=0 and DT=LT was evident at the 1 ms lead time ($p=0.017$). The difference between these two age groups was not significant, $F<1$. Although there was no significant overall effect of decay time in the oldest mice, there was a significant trend for the interaction between DT and LT favoring the abrupt transition at lead times of 5 and 10 ms ($p=0.16$).

2. Experiment 2

Figure 2 depicts the mean relative response amplitudes (±SEM) for the two noise decay times across lead time, at three noise levels. The lines, solid for DT=0 and broken for DT=LT, are the nonlinear regression lines passing through the point 1.0 at a lead time of 0 ms, and providing an exponential decay function to the data. Both the time constants and the asymptotic plateau for the relative response strength generally increased with the decrease in noise level. For decreasing noise level in sequence, the time constants for DT=0 were 2.33, 3.73, and 3.97 ms; for DT=LT, 2.32, 4.55, and 7.94 ms; for DT=0 the plateaus were 0.13, 0.16, and 0.44; and for DT=LT, 0.31, 2.87, and 0.44. Although the differences in noise level produced differences in the strength of inhibition and in its time course, the difference between the inhibitory effects of the abrupt versus the ramped offset was evident in each noise condition. The ANOVA of these relative data provided in Table II found significant main effect of level, lead time, decay time, and an interaction between level and lead time, but not level and decay time ($F=0.06$). The pattern of effects was the same in male and female mice. There was an overall significant difference between the abrupt and the ramped decay time at a lead time of 1 ms ($p<0.001$), and this difference did not interact with noise level ($F=0.44$).

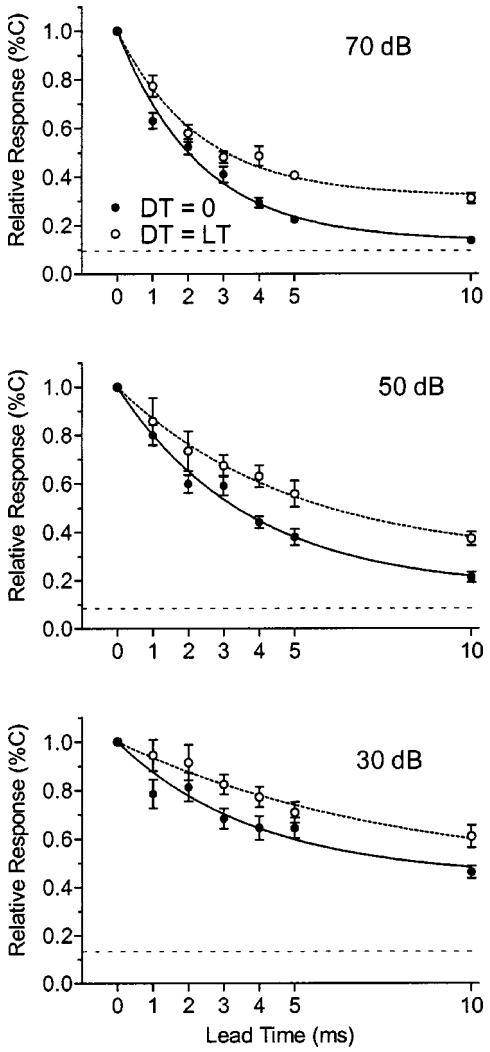


FIG. 2. Mean relative amplitude (±SEM) for the startle reflex when the eliciting stimulus was presented alone (the control level at LT=0) or when preceded by noise offset at the depicted lead times, this being the duration of the brief gap between noise offset and the onset of the startle stimulus. In one condition the decay time of the offset was abrupt (DT=0) and in the other condition the linear ramped decay time was the same duration as the lead time (DT=LT). The lines describe the fit to the data provided by an exponential decay function with two parameters, the time constant and the asymptotic plateau. The horizontal dashed line is the background activity level in the absence of any stimulus. The performance of adult mice, 5–6 months in age, is shown at three noise levels.

IV. DISCUSSION AND CONCLUSIONS

A. Changes in ASR amplitude with age and with noise level

Older mice responded less vigorously than younger mice, as has been previously reported in old mice (Parham and Willott, 1988; Ison *et al.*, 1998; Barsz *et al.*, 2002), in old rats (Krauter *et al.*, 1981), and in elderly human subjects (Ford *et al.*, 1995). About 30% of the variance in the ASR

amongst old CBA mice can be attributed to individual differences in hearing thresholds as measured with the ABR, but young mice with relatively poor hearing startle more vigorously than old mice with very good hearing for their age, the two groups thus having comparable ABR thresholds (Barsz *et al.*, 2002). One major additional cause of the decrement in startle amplitude in old humans as well as laboratory animals is likely to be the loss of muscle mass and changes in the composition and neural innervation of the skeletal muscles (Einsiedel and Luff, 1992). A second important factor may be changes in the levels of central neurotransmitters that have been shown to modulate startle reflex activity in young animals (Davis, 1980).

Startle behavior in young rats has been shown to be affected by background noise level (Hoffman and Fleshler, 1963; Ison and Hammond, 1971; Davis, 1974; and many others) As was found here, the effect is usually nonmonotonic, that is, there is an optimal level for facilitation beyond which the response is depressed rather than enhanced in its presence. The enhancement of startle by noise in the rat is diminished following administration of an anxiolytic drug (Kellogg *et al.*, 1991), suggesting that it is an arousal effect of the stimulating noise. The downturn in the amplitude of the response at high noise levels has been attributed to masking (Davis, 1974). In the present context the primary interest in this variable was its potential for diminishing the difference in the inhibitory effects of abrupt versus ramped offsets, as this would suggest that the effect of age seen in experiment 1 might have resulted because the effective level of the 70 dB noise carrier was diminished by hearing loss in the oldest mice. Both the overall level of inhibition and the time constant for the development of inhibition by abrupt noise offset were altered by variation in the background level, but this had little effect on the difference between ramped and abrupt offsets. In this respect these data agree with those presented for a single human listener by Miller (1948), though they were collected in a very different experimental paradigm, and, further, they indicate that the failure to find a difference in the behavioral effect of abrupt versus ramped noise offsets did not result from the modest hearing loss of the old CBA mouse.

B. The inhibitory effect of noise offset

The present results confirm previous findings (Stitt *et al.*, 1973; Ison, 1982; Ison *et al.*, 1998) that noise offset in rodents has a profound inhibitory effect on the acoustic startle reflex. The range of potential explanations of this effect is limited by the fact that the onset latency of the startle reflex is very fast, about 12 ms in indirect behavioral measures of applied force (Hoffman and Wible, 1970), 6 ms in EMG measures (Hammond *et al.*, 1972), or 3 ms in single unit activity in the motor brainstem (Carlson and Willott, 1998; Wu *et al.*, 1988). The observations here and in Ison *et al.* (1998) that significant inhibition is evident when noise offset precedes the acoustic startle stimulus by as little as 1 ms suggests that the stimulus must affect neural processing within the afferent pathways of the startle reflex, either directly by diminishing their sensitivity to afferent input, or indirectly, by engaging fast oligosynaptic feed-forward or feedback inhibitory circuits. In addition, the inhibitory phenomenon is clearly not a product of either the typical forward-masking mechanism or of sensory adaptation because it increases rather than diminishes in strength over time. It may be akin to the more complex form of forward masking observed in some cells of the dorsal cochlear nucleus by Kaltenbach *et al.* (1993), in which responsivity to a probe tone declined in strength over a period of 5 to 10 ms as it was increasingly delayed with respect to the prior masking tone. Kaltenbach *et al.* (1993) reported that this form of masking occurred only to the extent that there was a distinguishable gap between the probe and the masker. This is evident in the present data as well, as the noise by itself in the absence of the gap tended to enhance rather than depress the startle response.

As in a prior experiment (Ison *et al.*, 1998), the asymptotic inhibitory effect of the abrupt noise offset was systematically reduced with increasing age, which, however, did not affect the time constant for the growth of inhibition with the increase in the duration of the quiet period between noise offset and startle onset. A similar decrement in old CBA mice in the asymptotic inhibitory effect of a complete gap presented prior to the startle stimulus was described by Barsz *et al.* (2002), the gap in the case consisting of both the noise offset and the subsequent noise onset, with the startle occurring 60 ms later. Both of these findings are comfortably explained by assuming that the neural responses to noise transients, to either the abrupt noise decrement at the beginning of the gap in the present experiment and in Ison *et al.* (1998), or to the abrupt noise increment at the end of the gap in the experiments reported in Barsz *et al.* (2002), are diminished with age. Indeed, Barsz *et al.* (2002) as well as Walton *et al.* (1998) showed that the old CBA mouse has a reduced neural response in the inferior colliculus to noise onset at the end of the gap, and a slower recovery in this response with increasing suprathreshold gap durations. Given this physiological support for a diminished response to the end of the gap in the old mouse, there would seem no need to assume that in addition gap detection suffers because the old mouse is unable to follow the rapid changes in noise level at its beginning. However, the present data suggest that old age affects both the processes that are important to gap detection (Plomp, 1964), in showing that the behavioral effect of the difference between an abrupt and a ramped noise offset at the beginning of a gap emerges only at longer ramp times in the ''old-old'' mouse compared to younger mice.

One serious question is how a more slowly decaying excitatory response during the quiet period of the gap might be observed in the neurons in the inferior colliculus, because at this level of the brainstem the most sensitive gap thresholds of 1 to 2 ms appear in ''onset phasic'' cells that tend not to respond for continuous background noise. Thus, as these cells do not response to the noise prior to the gap, they could not be expected to show any greater persistence of firing during the gap. But Walton *et al.* (1998) also described a ''sustained'' cell type which does continue to respond during noise and reflects the presence of the gap by a reduction in its firing rate during the gap, rather than an onset response to noise at its end. In principle these cells could show such an

effect of greater persistence in the gap in old mice, and, in fact, consistent with the present data, it was observed that for these cells the average gap threshold that was just 4 ms in 2–3-month-old mice increased in 2-year-old mice to an average of 10 ms. This difference reveals the greater persistence of a level of neural activity that was appropriate for the presence of noise into the quiet gap for 10 ms in the old mice. Thus this aspect of the electrophysiological data obtained by Walton *et al.* (1998) agrees with the behavioral effect we see in the present experiment. Both of these effects indicate that behavioral gap direction in very old mice must suffer in part because of the relatively slow decline of afferent activity during the silent period of the gap.

Like young adult mice, young adult human listeners are also sensitive to the difference between abrupt noise offsets and ramped offsets of no more than 1 ms in duration (Allen *et al.*, 2002). It will now be especially of interest to determine whether elderly human listeners display the same loss of sensitivity to this variable of offset-ramp times as can be shown in old mice.

Our main experiment was based on the idea of von Bekesy (1933/1960) that the decay time of afferent activity after stimulus offset could be assessed by comparing the sensory impression of a sound with an abrupt offset with that of a ramped offset, to yield the longest ramp time that could not be distinguished from an abrupt offset. His analysis leads very easily to the gap detection method of determining the decay of auditory sensation as developed by Plomp (1964). The general theoretical importance of this latter measure of temporal acuity is shown in the observations that gap detection performance in human listeners is correlated with their speech perception (Tyler *et al.*, 1982; Snell and Frisina, 2000; Snell *et al.*, 2002). Past research has established that gap detection performance is diminished similarly in elderly human listeners and in old mice, as recently described in Barsz *et al.* (2002). The CBA mouse strain has served as an animal model for examining the effects of age on central auditory processing because it undergoes only modest cochlear pathology and associated hearing loss in old age (Willott, 1996), and because it shares with other animal models that capture significant attributes of human presbycusis the potential for understanding their underlying physiological substrate. The hypothesis that an age-related decline in the rate of decay in excitation at noise offset contributes to the loss of temporal acuity with age is consistent with the idea that this process is not passive, but depends in part on active inhibition (Bowen *et al.*, 2003). The rapid suppression of neural activity at stimulus offset may result in the mouse, and presumably other mammalian species, from a suppressive influence of inhibitory neural networks (Ferragamo *et al.*, 1998; Wickesberg and Oertel, 1990) or from the intrinsic membrane properties of certain auditory cells specialized for the rapid transmission of acoustic information (Bal and Oertel, 2001). Considerable research by Caspary and his colleagues has shown significant changes in the inhibitory neurotransmitter systems of the brainstem in the rat, a related animal model, in the inferior colliculus [e.g., Milbrandt *et al.* (1997) for GABA] and the cochlear nucleus [Krenning *et al.* (1998) for glycine]. At present the available data in the mouse are limited to a study of changes in glycine in the cochlear nucleus that reveal effects of age only in C57BL mice with severe hearing loss, not the CBA (Willott *et al.*, 1997). However, other work in the CBA mouse indicates that there are age-related differences in the expression of ion channels that are critical to fast temporal processing in both excitatory and inhibitory systems of the brain stem (Zettel and Frisina, 2003), and this, too, may contribute to changes in temporal processing with age. It is possible that a greater understanding of the neural bases of these documented age-related changes in auditory function in mice with minimal peripheral hearing loss will lead to novel pharmacological treatments for some aspects of presbycusis in the old mouse, and, potentially, for similar aspects of presbycusis in the elderly human.

ACKNOWLEDGMENTS

This research was supported by USPHS NIH-NIA Research Grant No. AG09524. A preliminary report of this work was presented at the 23rd Midwinter Research Meeting of the Association for Research in Otolaryngology, February, 2001. We thank Jason Castro, Manu Chopra, and Vincent Nalbone for their assistance in data collection.

Allen, P. D., Virag, T. M., and Ison, J. R. (**2002**). ''Humans detect gaps in broadband noise according to effective quiet time without additional cues from abrupt envelope changes,'' J. Acoust. Soc. Am. **112**, 2967–2974.

Bal, R., and Oertel, D. (**2001**). '' Potassium currents in octopus cells of the mammalian cochlear nucleus,'' J. Neurophysiol. **86**, 2299–2311.

Barsz, K., Ison, J. R., Snell, K. B., and Walton, J. P. (**2002**). ''Behavioral and neural measures of auditory temporal acuity in aging humans and mice,'' Neurobiol. Aging **23**, 565–578.

Bowen, P. G., Taylor, M. K., Lin, D., and Ison, J. R. (**2003**). ''Auditory cortex lesions impair both temporal acuity and intensity discrimination in the rat, suggesting a common mechanism for sensory processing,'' Cerebral. Cort. (in press).

Carlson, S., and Willott, J. F. (**1998**). ''Caudal pontine reticular formation of C57BL/6J mice: responses to startle stimuli, inhibition by tones, and plasticity,'' J. Neurophysiol. **79**, 2603–2614.

Davis, M. (**1974**). ''Sensitization of the rat startle response by noise,'' J. Comp. Physiol. Psychol. **87**, 571–581.

Davis, M. (**1980**). ''Neurochemical modulation of sensory-motor reactivity: acoustic and tactile startle reflexes,'' Neurosci. Biobehav. Rev. **4**, 241–263.

Einsiedel, L. J., and Luff, A. R. (**1992**). ''Alterations in the contractile properties of motor units within the aging rat medical gastrocnemius,'' J. Neurol. Sci. **112**, 170–177.

Ferragamo, M. J., Golding, N. L., and Oertel, D. (**1998**) ''Synaptic inputs of stellate cells in the ventral cochlear nucleus,'' J. Neurophysiol. **79**, 51–63.

Ford, J. M., Roth, W. T., Isaaks, B. G., White, P. M., Hood, S. H., and Pfefferbaum, A. (**1995**). ''Elderly men and women are less responsive to startling noises: N1, P3 and blink evidence,'' Biol. Psychol. **39**, 57–80.

Hammond, G. R., McAdam, D. W. and Ison, J. R. (**1972**). ''Effects of prestimulation on the electromyographic response associated with the acoustic startle reaction in rats,'' Physiol. Behav. **8**, 535–537.

Hoffman, H. S., and Fleshler, M. (**1963**). ''Startle reaction: Modification by background acoustic stimulation,'' Science (Washington, DC, U.S.) **141**, 928–930.

Hoffman, H. S., and Wible, B. L. (**1970**). ''Role of weak signals in acoustic startle,'' J. Acoust. Soc. Am. **47**, 489–497.

Ison, J. R. (**1982**). ''Temporal acuity in auditory function in the rat: reflex inhibition by brief gaps in noise,'' J. Comp. Physiol. Psychol. **96**, 945–954.

Ison, J. R., and Hammond, G. R. (**1971**). ''Modification of the startle reflex in the rat by changes in the auditory and visual environments,'' J. Comp. Physiol. Psychol. **75**, 435–452.

Ison, J. R., Agrawal, P., Pak, J., and Vaughn, W. J. (**1998**). ''Changes in temporal acuity with age and with hearing impairment in the mouse: a study of the acoustic startle reflex and its inhibition by brief decrements in noise level,'' J. Acoust. Soc. Am. **104**, 1696–1704.

Ison, J. R., Castro, J., Allen, P. D., Virag, T. M., and Walton, J. P. (**2002**). ''The relative detectability for mice of gaps having different ramp durations at their onset and offset boundaries,'' J. Acoust. Soc. Am. **112**, 740–747.

Kaltenbach, J. A., Meleca, R. J., Falzarano, P. R., Myers, S. F., and Simpson, T. H. (**1993**). ''Forward masking properties of neurons in the dorsal cochlear nucleus: possible role in the process of echo suppression,'' Hear. Res. **67**, 35–44.

Kellogg, C. K., Sullivan, A. T., Bitran, D., and Ison, J. R. (**1991**). ''Modulation of noise-potentiated acoustic startle via the benzodiazepine-gamma-aminobutyric acid receptor complex,'' Behav. Neurosci. **105**, 640–646.

Krauter, E. E., Wallace, J. E., and Campbell, B. A. (**1981**). ''Sensory-motor function in the aging rat,'' Behav. Neural Biol. **31**, 367–392.

Krenning, J., Hughes, L. F., Caspary, D. M., and Helfert, R. H. (**1998**). ''Age-related glycine receptor subunit changes in the cochlear nucleus of Fischer-344 rats,'' Laryngoscope **108**, 26–31.

Li, H. S., and Borg, E. (**1991**). ''Age-related loss of auditory sensitivity in two mouse genotypes,'' Acta Oto-Laryngol. **111**, 827–834.

Milbrandt, J. C., Hunter, C., and Caspary, D. M. (**1997**). ''Alterations of $GABA_A$ receptor subunit mRNA levels in the aging Fischer 344 rat inferior colliculus,'' J. Comp. Neurol. **379**, 455–465.

Miller, G. A. (**1948**). ''The perception of short bursts of noise,'' J. Acoust. Soc. Am. **20**, 160–170.

Parham, K., and Willott, J. F. (**1988**). ''Acoustic startle response in young and aging C57BL/6J and CBA/J mice,'' Behav. Neurosci. **102**, 881–886.

Plomp, R. (**1964**). ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.

Snell, K. B., and Frisina, D. R. (**2000**). ''Relationships among age-related differences in gap detection and word recognition,'' J. Acoust. Soc. Am. **107**, 1615–1626.

Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (**2002**). ''Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity,'' J. Acoust. Soc. Am. **112**, 720–727.

Stitt, C. L., Hoffman, H. S., and Marsh, R. (**1973**). ''Modification of the rat's startle reaction by termination of antecedent acoustic signals,'' J. Comp. Physiol. Psychol. **84**, 207–215.

Tyler, R. S., Summerfield, Q., Woods, E. J., and Fernandes, M. (**1982**). ''Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **72**, 740–752.

von Bekesy, G. (**1933/1960**). *Experiments in Hearing* (McGraw-Hill, New York).

Walton, J. P., Frisina, R. D., and O'Neill, W. E. (**1998**). ''Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse,'' J. Neurosci. **18**, 2764–2776.

Wickesberg, R. E., and Oertel, D. (**1990**). ''Delayed, frequency-specific inhibition in the cochlear nuclei of mice: a mechanism for monaural echo suppression,'' J. Neurosci. **10**, 1762–1768.

Willott, J. F. (**1996**). ''Anatomic and physiologic aging: a behavioral neurscience perspective,'' J. Am. Acad. Audiol. **7**, 141–151.

Willott, J. F., Milbrandt, J. C., Bross, L. S., and Caspary, D. M. (**1997**). ''Glycine immunoreactivity and receptor binding in the cochlear nucleus of C57BL/6J and CBA/CaJ mice: effects of cochlear impairment and aging,'' J. Comp. Neurol. **385**, 405–414.

Wu, M. F., Suzuki, S. S., and Siegel, J. M. (**1988**). ''Anatomical distribution and response patterns of reticular neurons active in relation to acoustic startle,'' Brain Res. **457**, 399–406.

Young, J. S., and Fechter, L. D. (**1983**). ''Reflex inhibition procedures for animal audiometry: A technique for assessing ototoxicity,'' J. Acoust. Soc. Am. **73**, 1686–1693.

Zettel, M. L., and Frisina, R. D. (**2003**). ''KV3.1 channel expression in the MNTB of old CBA mice,'' Assoc. Res. Otolaryngol. Abs. **26**, 50.

Whole-lung resonance in a bottlenose dolphin (*Tursiops truncatus*) and white whale (*Delphinapterus leucas*)

James J. Finneran
Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152-5001

(Received 12 July 2002; accepted for publication 18 February 2003)

An acoustic backscatter technique was used to estimate *in vivo* whole-lung resonant frequencies in a bottlenose dolphin (*Tursiops truncatus*) and white whale (*Delphinapterus leucas*). Subjects were trained to submerge and position themselves near an underwater sound projector and a receiving hydrophone. Acoustic pressure measurements were made near the thorax while the subject was insonified with pure tones at frequencies from 16 to 100 Hz. Whole-lung resonant frequencies were estimated by comparing pressures measured near the subject's thorax to those measured from the same location without the subject present. Experimentally measured resonant frequencies for the white whale and dolphin lungs were 30 and 36 Hz, respectively. These values were significantly higher than those predicted using a free-spherical air bubble model. Experimentally measured damping ratios and quality factors at resonance were 0.20 and 2.5, respectively, for the white whale, and 0.16 and 3.1, respectively, for the dolphin.
[DOI: 10.1121/1.1575747]

PACS numbers: 43.80.Nd, 43.80.Ev [WA]

I. INTRODUCTION

Exposure to underwater shock waves or high-intensity sound waves may cause injury to submerged terrestrial or marine animals. The injury sites are often concentrated in gas-filled structures within the animal (e.g., Yelverton *et al.*, 1973, 1975). These structures are much more compressible than the surrounding tissue and act as pressure-release surfaces, i.e., the gas/tissue interface may be the site of a large pressure gradient and surface displacement that may rupture tissues. Animals with extensive gas-filled tissues have been shown to be at higher risk than other animals. For example, fish possessing swimbladders have been shown to be more susceptible to the effects of underwater blasts than species without swimbladders (Yelverton *et al.*, 1975; Wiley *et al.*, 1981; Goertner *et al.*, 1994). Marine mammals may be particularly susceptible to underwater shock waves or high-intensity sound because they possess lungs, small gas bubbles within the bowels, and air-filled cavities in the sinuses and middle ear.

Quantitative data regarding acoustic or blast-related injuries in marine mammals are rare; therefore, most efforts to predict injuries have relied on extrapolations from terrestrial mammal data (e.g., Ketten, 1995; Department of the Navy, 2001a, 2001b). In the 1970s, the Lovelace Foundation performed a series of studies to determine the effects of underwater explosions on human divers by exposing submerged sheep, dogs, and monkeys to underwater blasts (Yelverton *et al.*, 1973). These studies found that hemorrhaging in and around the lungs was the principal source of injury following blast exposure. Injury to the gastrointestinal tract, attributed to excitation of radial oscillations in small, naturally occurring gas bubbles, was also found. Data from Yelverton *et al.* (1973) were extrapolated to marine mammals and became the basis for a model to predict safe ranges for sea mammals exposed to underwater explosions (Goertner, 1982). In this model the lungs were treated as an equivalent-volume spherical air bubble. Goertner (1982) used equations for the radial oscillation of a submerged free-spherical bubble and ''safe'' exposure levels from the Yelverton *et al.* (1973) data to predict safe ranges for various charge weight/depth/animal size combinations. The potential for low-frequency tones to induce injury in submerged rodents has also been investigated (Dalecki *et al.*, 1999; Cudahy *et al.*, 1999; Dalecki *et al.*, 2002). These studies found that the lungs and surrounding tissues were common injury sites and were most susceptible when the sound frequency was near the lung resonant frequency.

Environmental compliance documents have used the Goertner model to predict the effects of and establish mitigation ranges for explosive tests (e.g., Department of the Navy, 1998, 2001a). The free-spherical bubble model has also been used to estimate marine mammal lung resonant frequencies and thus the potential for tonal sound sources to cause injury in these animals (e.g., McMullen and McCarthy, 1998). Unfortunately, no direct measurements have been made to verify if (1) whole-lung radial oscillation in marine mammals may be modeled using equations for a free-spherical air bubble, or (2) the lung volume changes with depth as a free-spherical air bubble undergoing isothermal compression/expansion.

Figure 1 illustrates the structure of the lungs in a bottlenose dolphin (*Tursiops truncatus*). Figure 1(a) is a representation of the dolphin's internal anatomy indicating the orientation of the lungs within the thorax. Figures 1(b) and (c) are axial ''slices'' from a computed tomography (CT) scan obtained from a live 140-kg female bottlenose dolphin. The approximate locations of the scans in Figs. 1(b) and (c) are indicated in Fig. 1(a) by the vertical lines labeled ''B'' and ''C,'' respectively. Figures 1(b) and (c) illustrate the complex shape and structure of the lungs and the approximate cross-sectional size of the inflated lungs for a dolphin of this size.

Although marine mammal lungs are similar in many re-

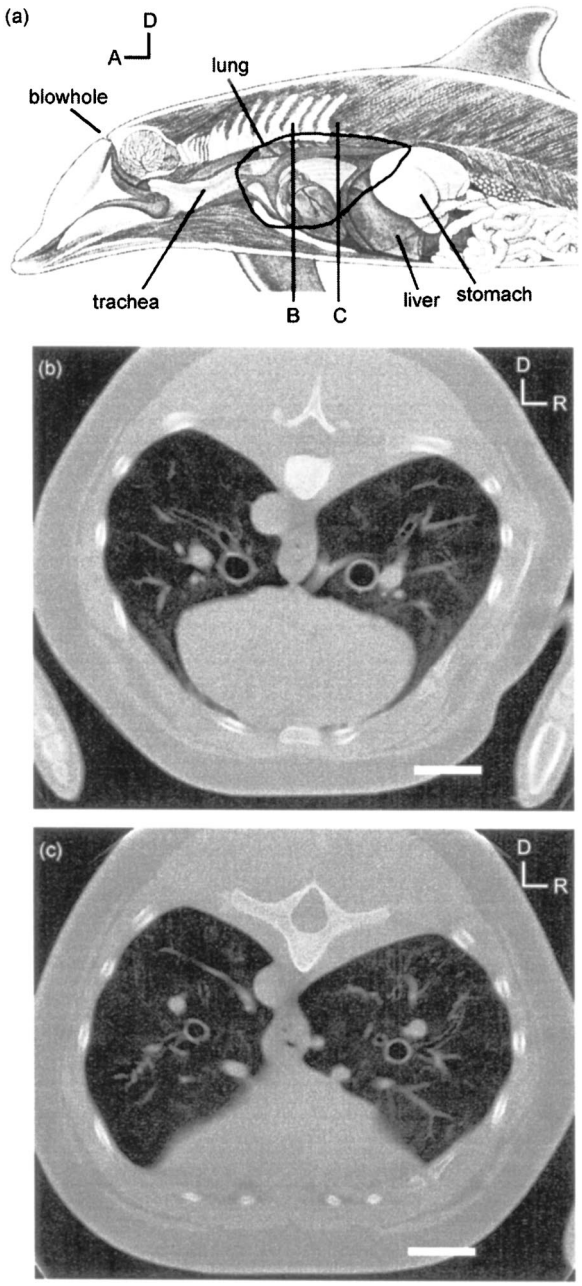


FIG. 1. (a) Lateral view of a bottlenose dolphin showing the location of the lungs within the thoracic cavity [the anatomical drawing by Barbara Stolen Irvine was modified from Ridgway (1972)]. Vertical bars (B,C) indicate the approximate locations of the transverse CT sections shown in (b) and (c), respectively. The CT scans were obtained with a live, 140-kg, 220-cm bottlenose dolphin. A—anterior, D—dorsal, R—right. The white horizontal bars represent 5 cm.

spects to those of terrestrial mammals, there are important anatomical differences between the two. Unlike terrestrial mammals, marine mammals have structural adaptations to their respiratory systems that allow them to withstand large variations in hydrostatic pressure and avoid ''chest squeeze'' and decompression sickness while diving and surfacing (Ridgway and Harrison, 1986). The chest cavity and lungs in odontocetes are highly elastic and resilient; many of the ribs are floating rather than being firmly attached to the sternum, allowing the rib cage and lungs to collapse at depth and expand when surfacing (Ridgway, 1986; Ridgway and Harrison, 1986; Rommel, 1990). Marine mammals also have cartilage and/or muscle surrounding the bronchi and bronchioles (Kooyman, 1973). Odontocetes have flexible, interlocking cartilage rings surrounding the trachea, bronchi, and bronchioles (Ridgway *et al.*, 1969). These arrangements allow the lungs, including the alveoli, to progressively collapse as the animal dives and prevent the bronchioles from prematurely collapsing and trapping air within the alveolar sacs (Kooyman, 1973). Such a mechanism would presumably prevent blood within the capillaries from coming into contact with nitrogen at high pressures, blocking diffusion of nitrogen into the blood below a certain depth (Kooyman, 1973). Lung collapse, including the alveoli, is estimated to be complete by 100-m depth in bottlenose dolphins (Ridgway and Howard, 1979).

The net effect of diving-related structural adaptations on the mechanical response of the lungs to underwater sound is unknown. It is likely that the free-spherical air bubble model underestimates the stiffness and damping found in the actual lung response, leading to underestimates of the resonant frequency and overestimates of the amplitude at resonance; however, there are no existing data regarding lung resonant frequencies in live marine mammals. This paper presents the results of a study to determine the feasibility of direct measurement of marine mammal whole-lung resonant frequencies using an acoustic backscatter technique. The experimental approach involved insonifying a trained subject and comparing the sound pressure measured near the lungs to the incident sound pressure alone (i.e., without the subject present). The measurement technique is presented, along with experimentally measured lung resonant frequencies for a bottlenose dolphin and a white whale (*Delphinapterus leucas*). The measured data are compared to theoretical predictions and analogous data from terrestrial mammals.

II. EXPERIMENTAL APPROACH

The acoustic backscatter technique has been previously used to measure fish swimbladder resonances (Batzler and Pickwell, 1970; McCartney and Stubbs, 1971) and whole-lung resonances in swine (Duykers and Percy, 1978), mice (Dalecki *et al.*, 1999), and humans (Martin *et al.*, 2000). Figure 2 illustrates the basic principles. An underwater sound source is used to generate tonal signals and insonify a target whose resonant frequency is desired. Two separate measurements are required. First, the sound pressure produced by the source is measured using a hydrophone positioned near the target [Fig. 2(a)]. This pressure is denoted as P_1 and consists of the sum of the incident pressure and the pressure scattered by the target. Next, the measurement is repeated without the target present [Fig. 2(b)]. The second measurement provides P_2, which consists of the incident pressure alone. The target resonant frequencies are determined by examining the complex ratio of P_1/P_2 at different frequencies.

A compressible structure, such as the lungs, whose size is small relative to the acoustic wavelength will undergo volume compression and expansion when insonified by a sound wave. This mode of oscillation is often referred to as the monopole or ''breathing'' mode, and may be modeled by treating the structure as a single degree-of-freedom mechani-

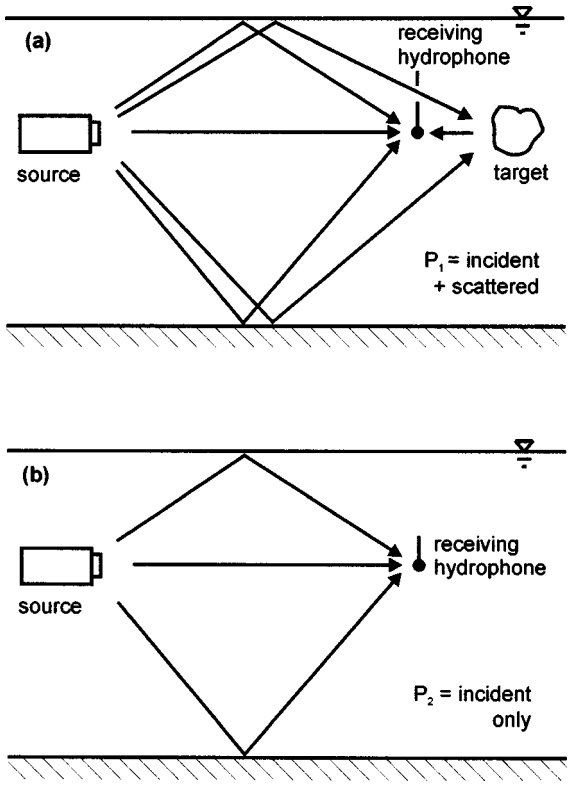


FIG. 2. The experimental backscatter technique relied on acoustic pressure measurements at a fixed location with (P_1) and without (P_2) the target present. Resonant frequencies are identified as the locations of local maxima in the amplitude of P_1/P_2, local minima in the imaginary part of P_1/P_2, and a maximum-to-minimum transition in the real part of P_1/P_2.

cal oscillator with mechanical stiffness k_s, mechanical damping b, and mass m (Devin, 1959). If this structure is insonified by an underwater sound source operating at a frequency f, the ratio P_1/P_2 is

$$\frac{P_1}{P_2}=1+\frac{K(f/f_n)^2}{1-(f/f_n)^2+j2\zeta(f/f_n)}, \tag{1}$$

where K is a (complex) constant, f_n is the undamped natural frequency (i.e., the resonant frequency) of the mechanical oscillator [$2\pi f_n=(k_s/m)^{1/2}$], ζ is the damping ratio [$\zeta=0.5b(k_s m)^{-1/2}$], and $j=(-1)^{1/2}$. The frequency at which damped free oscillations occur is the damped natural frequency, f_d. At frequencies near f_n the response of the structure is primarily controlled by the damping ratio, which may, in the general case, be a function of frequency. The amplitude at resonance and the sharpness of the amplitude peak at resonance are normally described using the damping ratio at resonance, ζ_n, or the quality factor Q[$Q=(2\zeta_n)^{-1}$]. For small values of ζ_n, $f_d \approx f_n$.

Figure 3 shows the theoretical value of P_1/P_2 as a function of frequency for a 10-cm-radius spherical air bubble in an unbounded body of water at atmospheric pressure. The bubble mass, stiffness, and damping were calculated from equations provided by Devin (1959) and Eller (1970). Surface tension, thermal damping, and viscous damping were neglected; at the frequencies of interest the viscous and thermal damping terms are negligible (Devin, 1959; Eller, 1970). The distance between the source and bubble center was 1 m;

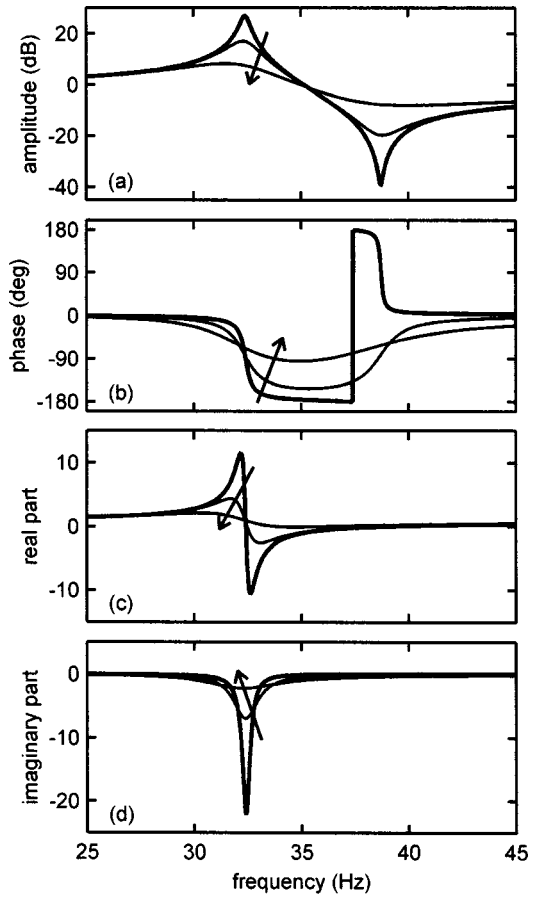


FIG. 3. Theoretical (a) amplitude; (b) phase; (c) real part; and (d) imaginary part of P_1/P_2 for a 10-cm free-spherical air bubble in water at 1 atm hydrostatic pressure. The thick line represents the bubble with the theoretical radiation damping only. The thin lines indicate 10-dB increments in the damping ratio. The arrows indicate the direction of increasing damping.

the pressures P_1 and P_2 were calculated at a distance of 75 cm from the source. Figures 3(a), (b), (c), and (d) show the amplitude, phase, real part, and imaginary part of P_1/P_2, respectively. The thick lines in Fig. 3 indicate the bubble response with the theoretical radiation damping only. The bubble resonance is revealed by the peak in the amplitude ratio, minimum value in the imaginary part of P_1/P_2, Im(P_1/P_2), and maximum-to-minimum transition in the real part of P_1/P_2, Re(P_1/P_2), located at approximately 32 Hz. The thin lines in Fig. 3 indicate the effects of increased damping. Each curve represents a 10-dB increase in the damping ratio, with the arrows indicating the direction of increasing damping. As the damping increases, the amplitudes of the maxima and minima are brought closer to zero, the peaks become more ''spread out,'' and the frequencies of the amplitude peak and imaginary part minimum are reduced.

III. METHODS

A. Experimental subjects

Experimental subjects consisted of a white whale (MUK, female, 32 years, approximately 540 kg, length 330 cm) and a bottlenose dolphin (NAY, male, 17 years, approximately 280 kg, 290 cm). Both subjects were healthy during

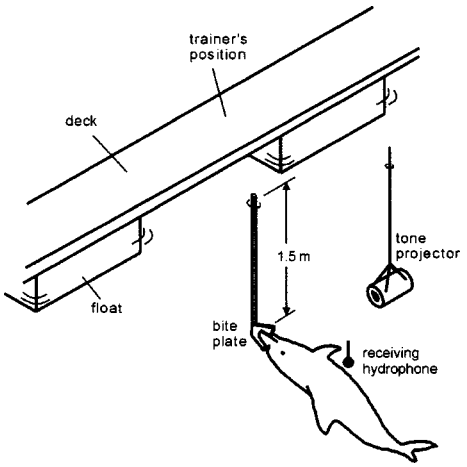


FIG. 4. Experimental setup for acoustic backscatter measurements with the dolphin and white whale.

the study as determined through periodic veterinary exams. Tests were conducted according to a protocol approved by the Institutional Animal Care and Use Committee at the Space and Naval Warfare Systems Center, San Diego, and followed all applicable U.S. Department of Defense guidelines.

B. Procedure

Figure 4 shows the experimental setup. Tests were conducted in a 12×12-m floating netted enclosure located in San Diego Bay. Water depth ranged from approximately 4–6 m. The apparatus consisted of a polyvinyl chloride frame containing a plastic ''biteplate'' at a depth of 1.5 m. Both subjects previously participated in a study to measure low-frequency hearing thresholds (see Finneran *et al.*, 2002) and were trained to position themselves on this apparatus. The hearing study required the subjects to position themselves on their left side (see Fig. 4); this orientation was maintained during the acoustic backscatter measurements. An underwater sound projector (USRD J13) was located ventral to the subject, approximately 1 m from the thorax. A receiving hydrophone (B&K 8105) was positioned between the projector and the subject, approximately 15 cm from the subject's thorax.

Measurements with the subjects were conducted over a series of ''dives.'' Each dive began with the trainer signaling the subject to position itself on the biteplate. While on the biteplate the subject was insonified by one or more single-frequency tones. Tones were generated using a National Instruments PCI-MIO-16E-1 multifunction board in a personal computer, filtered (Ithaco 4302), and amplified (Hafler P7000) before being input to the J13 sound projector. Tones were 1.5 s in duration with 250-ms rise and fall times. The acoustic pressure was measured during each tone presentation. The hydrophone output was amplified (B&K 2692) and filtered (Avens AD8DHL8B-1K-256K) before being digitized (PCI-MIO-16E-1) at 6 kHz. The digitized pressure waveforms were weighted using a flat-top window; frequency spectra were obtained using a fast Fourier transform (FFT). The pressure amplitude and phase were obtained from the frequency spectra; the pressure phase angle was measured relative to the projector input voltage phase. The dive was concluded when the trainer sounded an underwater buzzer to signal the subject to return to the surface for fish reward. Most dives were 1–2 min in duration. After collecting data from 16 to 100 Hz (in 2-Hz increments), the subject was removed from the test enclosure and the pressure amplitude and phase measurements repeated at each frequency to obtain P_2.

C. Data analysis

The damped natural frequency may be estimated from experimental measurements of P_1/P_2 in one of three ways: (1) the frequency where $|P_1/P_2|$ has a maximum; (2) the frequency where $\mathrm{Re}(P_1/P_2)$ crosses one with a maximum and minimum at higher and lower frequencies; or (3) the frequency where $\mathrm{Im}(P_1/P_2)$ has a minimum. Method (3) is generally thought to give the most reliable results under all conditions; however, locating the actual frequency where a minimum exists may be difficult if the frequency resolution is poor and/or the peak is sharp. In this study, f_d was estimated using the mean value from methods (2) and (3). The undamped natural frequency f_n was estimated from f_d and ζ_n using

$$f_d = f_n\sqrt{1-\zeta_n^2}. \tag{2}$$

The damping ratio at resonance is often estimated by halving the ratio of the −3-dB frequency bandwidth to the peak frequency; however, this technique may not be suitable to systems with high damping ratios. Two additional methods to estimate ζ_n, using the real and imaginary parts, are given by

$$\zeta_n = \frac{1}{2}\frac{(f_B/f_C)^2-1}{(f_B/f_C)^2+1}, \tag{3}$$

where f_B and f_C are the frequencies on either side of f_n where $\mathrm{Re}(P_1/P_2)$ reaches a local maximum or minimum $(f_B > f_C)$, and

$$\zeta_n = \frac{|f_D - f_E|}{2f_d}, \tag{4}$$

where f_D and f_E are the frequencies where $\mathrm{Im}(P_1/P_2)$ is equal to one-half the value of $\mathrm{Im}(P_1/P_2)$ at f_d (Chu and Wang, 1980). Equations (3) and (4) will not generally produce the same result. In this study, ζ_n was estimated using the mean of the results of Eqs. (3) and (4).

IV. RESULTS

Figures 5(a), (b), (c), and (d) show the amplitude, phase, real part, and imaginary part of P_1/P_2, respectively, as functions of frequency for the white whale, MUK (thick line). Also shown for comparison is the ratio of two independent pressure measurements made without the subject present (thin line). The P_1/P_2 ratio exhibits substantial differences from the ''ambient-only'' measurements performed without the subject present, specifically a maximum/minimum pair in the amplitude and real part and a minimum in the phase and imaginary part. The experimentally measured P_1/P_2 ratio is

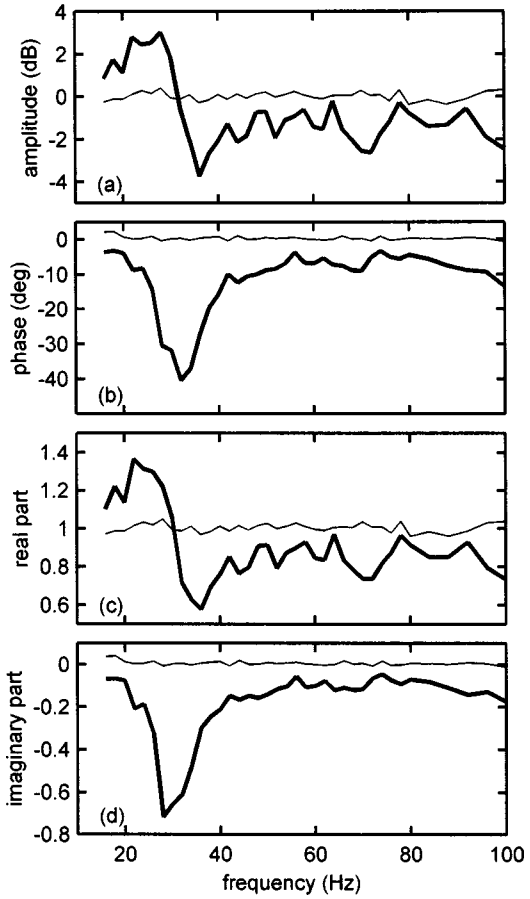


FIG. 5. Ratio of P_1/P_2 measured with a white whale subject and the configuration shown in Fig. 4. The whole-lung resonance exists at 30 Hz.

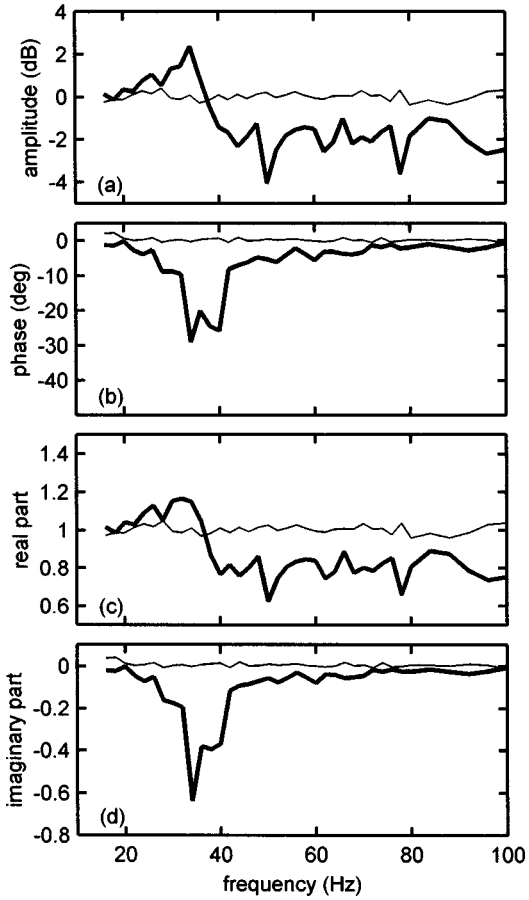


FIG. 6. Ratio of P_1/P_2 measured with a bottlenose dolphin subject and the configuration shown in Fig. 4. The whole-lung resonance exists at 36 Hz.

similar to that predicted from the spherical bubble model; however, the relatively low amplitude and broad peaks of the experimentally measured P_1/P_2 ratio reveal that the damping is substantially larger than that predicted from the spherical bubble model. The damped natural frequency was estimated to be 29 Hz. Mean values for the damping ratio at resonance and quality factor were 0.20 and 2.5, respectively. The undamped natural (resonant) frequency calculated from f_d, ζ_n, and Eq. (2) was 30 Hz.

Figure 6 is analogous to Fig. 5 and shows the results obtained for the dolphin, NAY. As with the white whale MUK, the P_1/P_2 ratio exhibited a maximum/minimum pair in the amplitude and real part and a minimum in the phase and imaginary part; these features were not present in the ambient-only measurement. The damped natural frequency for NAY was 35 Hz. Mean values for ζ_n and Q were 0.16 and 3.1, respectively. The resonant frequency was estimated to be 36 Hz.

Optimization techniques are often used to fit model parameters (e.g., f_n and ζ_n) to experimentally measured data. However, in the present case this approach is complicated by the factor K in Eq. (1), which requires knowledge of the lung size and distance from source to lung center, and the unknown frequency dependency of the mechanical damping. Nonetheless, to verify the experimental estimates of f_n and ζ_n, a least-squares technique was used to fit Eq. (1) to the experimental data. For this procedure the lungs were represented by spheres of radii 18 and 14 cm for MUK and NAY, respectively. The distance from source to lung center was estimated as 1.30 and 1.25 m for MUK and NAY, respectively. Radiation damping was assumed, resulting in the damping being a function of f^2. The optimization process yielded resonant frequencies of 30 and 36 Hz for MUK and NAY, respectively. These values are identical to those estimated directly from the experimental data. Damping ratios at resonance from the optimization process were 0.20 and 0.25 for MUK and NAY, respectively. The value for MUK was identical to that estimated using Eqs. (3) and (4); however, the damping ratio for NAY was substantially higher than that estimated from the experimental data. Resonant frequencies produced from the optimization process were not strongly dependent on the assumed value of K; however, the damping ratios were highly dependent on K; therefore, the value of ζ_n estimated directly from the measured P_1/P_2 ratio (0.16) was considered to be the more reliable estimate.

V. DISCUSSION

The objective of this study was to determine if the acoustic backscatter technique could be used to estimate whole-lung resonant frequencies in odontocetes. The data shown in Figs. 5 and 6 confirm that odontocete lungs undergo a breathing mode oscillation when insonified at low frequencies and that the resonant frequency of this mode may be estimated using the backscatter technique. The esti-

mated values for the resonant frequencies in a 280-kg bottlenose dolphin and a 340-kg white whale were 36 and 30 Hz, respectively, at 1.5-m depth.

The experimental technique assumed that any differences observed between the pressures P_1 and P_2 were caused by the presence of the subject. This was an approximation which may have led to measurement errors if there were reflections from underwater objects that were blocked by the subject but reached the hydrophone when the subject was not present. The measurement technique also assumed that the pressure was constant across the surface of the lungs. This was likely not the case because of the relatively short distance between the subject and sound projector. To avoid these two problems, future backscatter measurements should be made in deeper water using a larger source distance, if possible.

Ridgway *et al.* (1969) reported lung volume-to-body mass ratios of 51 and 55 ml/kg for 138- and 200-kg Atlantic bottlenose dolphin (*T.t.* truncatus), respectively. Kooyman and Cornell (1981) reported a lung volume-to-body mass ratio of 91 ml/kg for a 285-kg Pacific bottlenose dolphin (*T.t.* gilli) [Pacific bottlenose dolphins are generally larger and deeper divers than their Atlantic counterparts (Ridgway, 1986)]. Using a range of 50–90 ml/kg as an estimate, lung volumes of 27–49 and 14–25 L were calculated for MUK and NAY, respectively. Using an equal-volume free-spherical bubble model, resonant frequencies at 1.5-m depth for MUK and NAY were calculated to be 16–20 and 20–24 Hz, respectively. These values are substantially lower than the experimentally measured values, most likely because the free-bubble model does not include the stiffness of the tissues that support or surround the lungs.

Models for fish swimbladders often include additional terms to account for the stiffness and damping of the surrounding tissues (Andreeva, 1964; McCartney and Stubbs, 1971; Love, 1978; Feuillade and Nero, 1998). Lung resonant frequencies measured from human divers in a hyperbaric chamber have been fit well using the spherical bubble model with an additional depth-independent stiffness (Martin *et al.*, 2000). The data resulting from the present study demonstrate that models for marine mammal lungs should also include additional stiffness and damping terms (i.e., beyond those of a free bubble).

Dalecki *et al.* (1999) reported mean whole-lung resonant frequencies (at 1 atm) of 325 and 420 Hz for adult (30 g) and young (10 g) mice and a mean resonant frequency of 175 Hz for 320-g rats. Cudahy *et al.* (1999) reported that these data were fit well by a power function with the form $f_n = 132M^{-1/4}$, where f_n is the resonant frequency (Hz) and M is the body mass (kg). Figure 7 compares this function (solid line) and the mice/rat data from Dalecki *et al.* (1999) with the experimentally measured values for the white whale and dolphin. The experimentally measured white whale and dolphin values (30 and 36 Hz, respectively) are approximately 10% higher than the power function predictions (27 and 32 Hz, respectively). Some discrepancy may be the result of testing at different depths—one would expect the whale and dolphin resonant frequencies to decrease with decreasing hydrostatic pressure and thus approach the terrestrial mammal

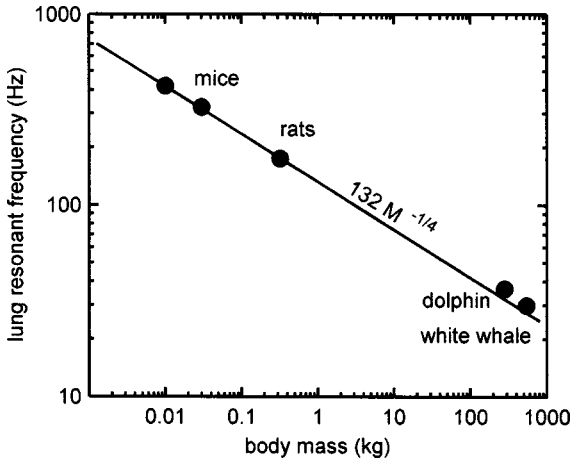


FIG. 7. Experimentally measured resonant frequencies from the dolphin and white whale compared to those reported by Dalecki *et al.* (1999) for young (10 g) and adult (30 g) mice and (320 g) rats. The solid line was derived from the terrestrial mammal data and has the form $f_n = 132M^{-1/4}$, where f_n is the resonant frequency (Hz) and M is the body mass (kg) (Cudahy *et al.*, 1999).

curve. Although the terrestrial mammal curve is reasonably close to the marine mammal data at shallow depth, it is unlikely that such an equation will apply to other depths as the lungs of marine mammals progressively collapse during diving. Backscatter measurements with trained marine mammals at various depths would allow equations to be developed for the variation of resonant frequency with depth in these animals.

A single degree of freedom lung model requires both the resonant frequency and damping ratio (or quality factor Q) to predict the actual response of the lung at resonance. Estimated damping ratios for MUK and NAY were 0.20 and 0.16. Estimated quality factors of 2.5 and 3.1 for MUK and NAY, respectively, are close to values of 5–7 reported for human lungs (Martin *et al.*, 2000) and 3–5 reported for goldfish swimbladders (Batzler and Pickwell, 1970), and much lower than those predicted by the free-spherical bubble model.

VI. CONCLUSIONS

The feasibility of an acoustic backscatter technique to measure whole-lung resonant frequencies in marine mammals was demonstrated. Estimated lung resonant frequencies were 30 and 36 Hz for a 540-kg white whale and a 280-kg dolphin, respectively. Estimated damping ratios and quality factors at resonance were 0.20 and 2.5, respectively, for the white whale, and 0.16 and 3.1, respectively, for the dolphin. Measured resonant frequencies and damping ratios were significantly higher than those predicted from a free-spherical bubble model. Measured resonant frequencies from this study at shallow depth were approximately 10% higher than predictions from a quarter-power scaling law derived from terrestrial mammal data.

ACKNOWLEDGMENTS

Subjects were trained by Jennifer Jeffress. Don Carder provided technical assistance. Dr. Sam Ridgway and Dr.

Dorian Houser provided valuable comments on the manuscript and, with Dr. William Van Bonn, provided the CT scan images. I thank Dr. John Sigurdson and two anonymous reviewers for critically reading the manuscript and providing helpful comments. Financial support was provided by the U.S. Office of Naval Research Marine Mammal Science and Technology Program, the U.S. Navy CNO (N45), and the SSC San Diego In-house Laboratory Independent Research (ILIR) program.

Andreeva, I. B. (**1964**). ''Scattering of sound by air bladders of fish in deep sound-scattering ocean layers,'' Sov. Phys. Acoust. **10**, 17–20.

Batzler, W. E., and Pickwell, G. V. (**1970**). ''Resonant acoustic scattering from gas-bladder fishes,'' Proceedings of the Int. Symp. Biological Sound Scattering in the Ocean, Maury Center for Ocean Science. 31 March–2 April 1970.

Chu, F. H., and Wang, B. P. (**1980**). ''Experimental determination of damping in materials and structures,'' in *Damping Applications for Vibration Control*, edited by P. J. Torvik, American Society of Mechanical Engineers AMD—Vol. 38 (American Society of Mechanical Engineers, New York), pp. 113–122.

Cudahy, E., Hanson, E., and Fothergill, D. (**1999**). ''Summary on the bioeffects of low frequency water borne sound,'' in Technical Report 3, Environmental Impact Statement for Surveillance Towed Array Sensor System Low-frequency Active (SURTASS LFA) Sonar.

Dalecki, D., Child, S. Z., and Raeman, C. H. (**2002**). ''Lung damage from exposure to low-frequency underwater sound,'' J. Acoust. Soc. Am. **111**, 2462(A).

Dalecki, D., Raeman, C. H., Child, S. Z., McAleavey, S. A., and Carstensen, E. L. (**1999**). ''Lung response to low-frequency sound,'' J. Acoust. Soc. Am. **106**, 2165(A).

Department of the Navy (**1998**). ''Final Environmental Impact Statement: Shock Testing the SEAWOLF submarine.''

Department of the Navy (**2001a**). ''Final Environmental Impact Statement: Shock Trial of the WINSTON S. CHURCHILL (DDG81).''

Department of the Navy (**2001b**). ''Final Environmental Impact Statement for Surveillance Towed Array Sensor System Low-frequency Active (SURTASS LFA) Sonar.''

Devin, Jr., C. (**1959**). ''Survey of thermal, radiation, and viscous damping of pulsating air bubbles in water,'' J. Acoust. Soc. Am. **31**, 1654–1667.

Duykers, L. R. B., and Percy, J. L. (**1978**). ''Low-frequency response of the submerged human lung,'' J. Acoust. Soc. Am. Suppl. 1 **64**, S97.

Eller, A. I. (**1970**). ''Damping constants of pulsating bubbles,'' J. Acoust. Soc. Am. **47**, 1469–1470.

Feuillade, C., and Nero, R. W. (**1998**). ''A viscous-elastic swimbladder model for describing enhanced-frequency resonance scattering from fish,'' J. Acoust. Soc. Am. **103**, 3245–3255.

Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2002**). ''Low-frequency acoustic pressure, velocity, and intensity thresholds in a bottlenose dolphin (*Tursiops truncatus*) and white whale (*Delphinapterus leucas*),'' J. Acoust. Soc. Am. **111**, 447–456.

Goertner, J. F. (**1982**). ''Prediction of underwater explosion safe ranges for sea mammals,'' NSWC TR 82-188 (Naval Surface Weapons Center, Dahlgren Division, White Oak Detachment, Silver Spring, MD).

Goertner, J. F., Wiley, M. L., Young, G. A., and McDonald, W. W. (**1994**). ''Effects of underwater explosions on fish without swimbladders,'' NSWC TR 88-114 (Naval Surface Warfare Center, Silver Spring, MD).

Ketten, D. R. (**1995**). ''Estimates of blast injury and acoustic trauma zones for marine mammals from underwater explosions,'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Netherlands), pp. 391–407.

Kooyman, G. L. (**1973**). ''Respiratory adaptations in marine mammals,'' Am. Zool. **13**, 457–468.

Kooyman, G. L., and Cornell, L. H. (**1981**). ''Flow properties of expiration and inspiration in a trained bottlenosed porpoise,'' Physiol. Zool. **54**, 55–61.

Love, R. H. (**1978**). ''Resonant acoustic scattering by swimbladder-bearing fish,'' J. Acoust. Soc. Am. **64**, 571–580.

Martin, J. S., Rogers, P. H., Cudahy, E. A., and Hanson, E. L. (**2000**). ''Low-frequency response of the submerged human lung,'' J. Acoust. Soc. Am. **107**, 2813(A).

McCartney, B. S., and Stubbs, A. R. (**1971**). ''Measurement of the acoustic target strengths of fish in dorsal aspect, including swimbladder resonance,'' J. Sound Vib. **15**, 397–420.

McMullen, F., and McCarthy, E. (**1998**). ''Acoustic Analysis of SWAC 4 Phase II,'' in SACLANTCEN Bioacoustics Panel Summary Record and Report, Annex H, SACLANTCEN Bioacoustics Panel, 15–16 June 1998.

Ridgway, S. H. (**1972**). *Mammals of the Sea* (Thomas, Springfield, IL).

Ridgway, S. H. (**1986**). ''Diving by cetaceans,'' in *Diving in Animals and Man*, edited by A. O. Brubakk, J. W. Kanwisher, and G. Sundness (Royal Norwegian Society of Science and Letters, Trondheim, Norway), pp. 33–62.

Ridgway, S. H., and Harrison, R. J. (**1986**). ''Diving dolphins,'' in *Research on Dolphins*, edited by M. M. Bryden and R. J. Harrison (Oxford University Press, Oxford), pp. 33–58.

Ridgway, S. H., and Howard, R. (**1979**). ''Dolphin lung collapse and intramuscular circulation during free diving: Evidence from nitrogen washout,'' Science (Washington, DC, U.S.) **206**, 1182–1183.

Ridgway, S. H., Scronce, B. L., and Kanwisher, J. (**1969**). ''Respiration and deep diving in a bottlenose porpoise,'' Science (Washington, DC, U.S.) **166**, 1651–1654.

Rommel, S. A. (**1990**). ''Osteology of the bottlenose dolphin,'' in *The Bottlenose Dolphin*, edited by S. Letherwood and R. R. Reeves (Academic, San Diego), pp. 29–49.

Wiley, M. L., Gaspin, J. B., and Goertner, J. F. (**1981**). ''Effects of underwater explosions on fish with a dynamical model to predict fishkill,'' Ocean Sci. Eng. **6**, 223–284.

Yelverton, J. T., Richmond, D. R., Fletcher, E. R., and Jones, R. K. (**1973**). ''Safe distances from underwater explosions for mammals and birds,'' prepared for the Defense Nuclear Agency by the Lovelace Foundation for Medical Education and Research, Report Number DNA 3114T.

Yelverton, J. T., Richmond, D. R., Hicks, W., Saunders, K., and Fletcher, E. R. (**1975**). ''The relationship between fish size and their response to underwater blast,'' prepared for the Defense Nuclear Agency by the Lovelace Foundation for Medical Education and Research, Report Number DNA 3677T.

Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear wave speeds in muscles

Jean-Luc Gennisson,[a] Stefan Catheline, Sana Chaffaï, and Mathias Fink
Laboratoire Ondes et Acoustique, E.S.P.C.I., Université Paris VII, U.M.R. 7587 C.N.R.S. 1503, 10 rue Vauquelin, 75231 Paris cedex 05, France

(Received 23 September 2002; revised 30 March 2003; accepted 14 April 2003)

From the measurement of a low frequency (50–150 Hz) shear wave speed, transient elastography evaluates the Young's modulus in isotropic soft tissues. In this paper, it is shown that a rod source can generate a low frequency polarized shear strain waves. Consequently this technique allows to study anisotropic medium such as muscle. The evidence of the polarization of low frequency shear strain waves is supported by both numeric simulations and experiments. The numeric simulations are based on theoretical Green's functions in isotropic and anisotropic media (hexagonal system). The experiments *in vitro* led on beef muscle proves the pertinent of this simple anisotropic pattern. Results *in vivo* on man biceps shows the existence of slow and fast shear waves as predicted by theory. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1579008]

PACS numbers: 43.80.Qf, 43.80.Vj, 43.80.Ev [FD]

I. INTRODUCTION

Anisotropy has long been studied in geophysics[1,2] or in nondestructive testing with ultrasound.[3,4] In biological media *ex vivo*, Hoffmeister[5] and Kuo[6] on tendon, Yoon[7] on bone or Levinson[8] and Andersen[9] on muscle quantify the anisotropic elastic moduli. Now elastographic techniques, using the propagation of a low frequency shear wave[10,11] can measure the Young's modulus of soft tissues. In this paper, we describe how these methods can be adapted to study the anisotropic shear moduli of simple muscles *in vitro* and *in vivo*. In the first part, the transient elastography technique[12,13] is presented. In the second part, theoretical considerations from a hexagonal anisotropic medium (transverse isotropic system) predict the existence of a slow and a fast shear wave. It is shown that the use of a rod as a low frequency wave source polarizes the shear strain wave on the first 50 mm and thus enables one to select either of the shear wave. These results follows numerical simulations based on theoretical Green's function in isotropic and transversely isotropic medium. In the third section we present some experimental results *ex vivo* on beef muscle and *in vivo* on human biceps.

II. TRANSIENT ELASTOGRAPHY

A. Experimental setup

An ultrasonic transducer (5 MHz) is applied at the surface of a homogeneous Agar-gelatin phantom (3% Agar and 5% gelatin) (Fig. 1). The transducer, a 7 mm diameter and a 35 mm focal depth, is set up on a vibrator (Brüel & Kjær, type 4810). An ultrasonic pulse echo system is used with a 2 kHz recurrence frequency. The ultrasonic signals are sampled at 50 MHz and stored using a 9-bit digitizer with 2 Mbytes memory. The low frequency pulse (50 to 150 Hz) is obtained with a conventional function generator. The basic idea is that the low frequency vibrations generated by a piston source (in this case the transducer itself) induce a low frequency motion of the scatterers inside the media. This motion can be detected and measured with a cross correlation algorithm on successive *A*-scans. Such a method gives the displacement along the axis of the ultrasound beam with an accuracy of 1 μm.

B. Displacement induced by an acoustic pulse

When a pulse is applied on a semi-infinite, isotropic and homogeneous solid, two kinds of waves are generated: a compressional and a shear waves. Typical speeds in soft tissues are, respectively, $V_p = 1500\ \mathrm{m\,s^{-1}}$ and $V_s = 5\ \mathrm{m\,s^{-1}}$. These speeds are connected to the elastic Lamé coefficients λ, μ in isotropic elastic media,

$$V_p = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \tag{1}$$

$$V_s = \sqrt{\frac{\mu}{\rho}}. \tag{2}$$

ρ is the density. In order to describe soft tissues elasticity, the usual parameter is the Young's modulus E,

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}. \tag{3}$$

In soft media λ is 10^6 times bigger than μ, thus a good approximation of E is

$$E \cong 3\mu. \tag{4}$$

Finally, we can write the Young's modulus as function of the shear wave speed

$$E = 3\rho V_s^2. \tag{5}$$

The measurement of the speed is deduced from the experimental displacements of Fig. 2 using a spectral analysis

[a]Electronic mail: jeanluc.gennisson@loa.espci.fr

 0001-4966/2003/114(1)/536/6/$19.00

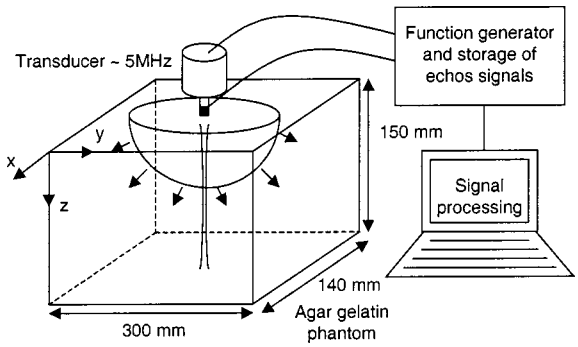


FIG. 1. Experimental setup. A 5 MHz ultrasonic transducer is set up on a vibrator. The acoustic impulse (100 Hz) is generated by the front face of the transducer while it works as a pulse echo system. The displacement field is calculated with a cross correlation technique on successive backscattered signals stored in memory.

at the central frequency. A linear fit on the phase as function of depth gives an accurate estimation of the phase velocity. The error is obtained from the standard deviation of the measurements from the linear fit. The values are $V_s = 2.88 \pm 0.03\ \mathrm{m\,s^{-1}}$ and using Eq. (5), $E = 8.29 \pm 0.14$ kPa (the value of ρ is found in the literature to be $1100\ \mathrm{kg\,m^{-3}}$). Thus one can determine precisely the Young's modulus in isotropic media. Now if we consider anisotropic media such as muscles, what kind of elasticity does this technique give? To answer this question we will use the theory of anisotropic elastic media.

III. THEORY OF ELASTIC WAVES IN TRANVERSE ISOTROPIC SOLID

A. General case

The muscle model considered in this paper presents a random distribution of fibers oriented in the same direction Fig. 3. This consideration entails the existence of a symmetry axis along the fibers. It is proven[3] that this kind of symmetry corresponds to a hexagonal system (transverse isotropy). The Christoffel's matrix c_{ijkl} of such a system contains 5 independent elastic constants. The eigenvectors of the Christoffel's tensor (implied in the wave equation) are associated to eigenvalues which are the speeds of waves in all directions.

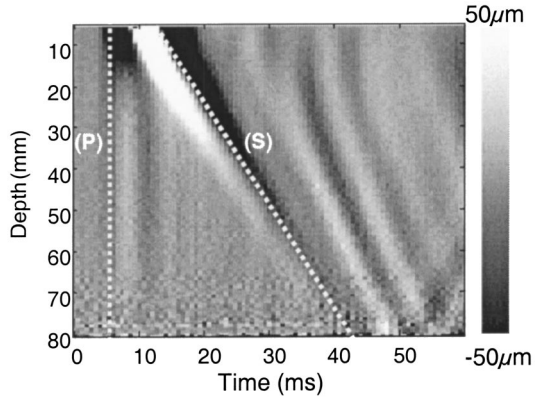


FIG. 2. Experimental displacement field obtained for an acoustic pulse (100 Hz). One can observe the displacements of a compressional wave (P) and of a shear wave (S) in an Agar-gelatin based phantom.

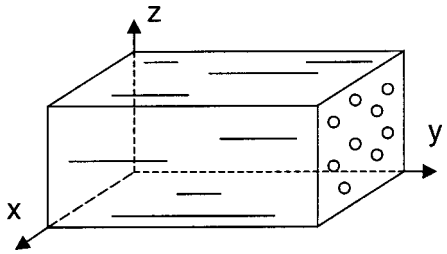


FIG. 3. Simple model of muscle. The fibers have the same orientation (along the y axis) and are randomly distributed on the plane (xz). The equivalent crystallographic model is the hexagonal system (transverse isotropy).

Nevertheless, we are only interested in wave propagation along the axis perpendicular to the muscle fibers (along x or z) since the direction parallel to the fibers is not easily accessible *in vivo* for the specific setup considered in this analysis. Moreover the weak backscattered signal for an ultrasonic beam parallel to this latter direction makes the use of transient elastography difficult. Therefore, in the direction perpendicular to the fibers (z direction) the speeds of the three different waves are

$$V_p = \sqrt{\frac{c_{11}}{\rho}}, \tag{6}$$

for the compressional wave,

$$V_s^{\perp} = \sqrt{\frac{c_{66}}{\rho}}, \tag{7}$$

for the shear wave with a polarization perpendicular to the fibers,

$$V_s^{\parallel} = \sqrt{\frac{c_{44}}{\rho}}, \tag{8}$$

for the shear wave with a polarization parallel to the fibers.

It clearly appears in Eqs. (7) and (8) that the elastic moduli c_{44} and c_{66} can be obtained if one is able to measure the speed of polarized shear waves. So how can one control the polarization of the low frequency shear wave along the axis of a piston source?

B. The polarization of shear strain waves

In the simple case of an isotropic medium, since the displacements preserve the symmetry of the initial boundary conditions, the shear strain field induced by a piston source is axial symmetric. Now the use of a rod source break this symmetry. Then the shear wave induces a shear strain field in a preferential direction (perpendicularly to the rod). We have first verify this hypothesis theoretically with the calculation of the exact Green's functions in an isotropic semi-infinite solid given by Gakenheimer and Miklowitz.[14] The parameters of the numerical simulation are $\rho = 1100\ \mathrm{kg\,m^{-3}}$, $V_p = 1500\ \mathrm{m\,s^{-1}}$, $V_s = 5\ \mathrm{m\,s^{-1}}$, a sampling frequency of 2 kHz, and a 80 mm rod length. Considering the rod as a 1 mm grid of point sources, we obtain the impulse response by summing the Green's function of all these individual point sources. This summation is the expression of the Rayleigh–Sommerfeld integral. Finally, a simple convolution of this

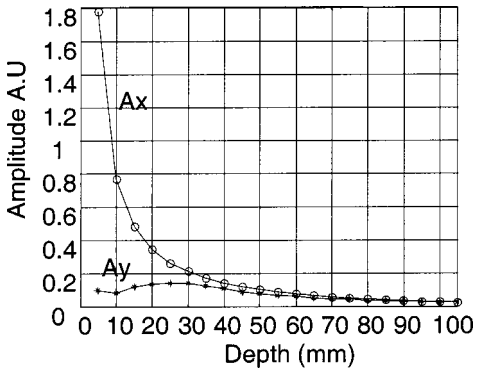


FIG. 4. Amplitude of the displacement gradient in the direction parallel (Ay) and perpendicular (Ax) to the rod versus depth. This simulation shows that the rod as a shear wave source favors the shear strain perpendicular to its main axis at least within the first centimeters of the isotropic solid model.

impulse response with the source excitation (a one-cycle sinusoid with a 100 Hz central frequency) gives the theoretical displacement field u_z. Figure 4 represents the amplitude of the gradient of the displacement (the shear strain field) in the direction parallel ($Ay=|\partial u_y/\partial y|$) and perpendicular ($Ax=|\partial u_x/\partial x|$) to the rod as function of depth. The axes are defined in Fig. 5. The strain field is smaller along the y axis than along the x axis until 40 mm depth. If we transpose this result to an anisotropic medium such as muscle, we can make the assumption that any speed measurement within this distance gives the speed of the shear wave with a polarization perpendicular to the fibers ($V_s^{\perp}$) when the rod is parallel to the fibers. Equally, if the rod is set perpendicular to the fibers, the speed of the shear wave with a polarization parallel to the fibers ($V_s^{\parallel}$) is obtained. We used a second numerical simulation in order to verify this latter assumption. The theoretical Green's function in an infinite transverse isotropic media are calculated by Vavryčuk[15] (see the Appendix). As shown by Sandrin,[16] Green's functions in isotropic infinite or semi-infinite soft media are very similar along the axis of the source. Thus the same assumption is made for anisotropic media. Again, by summing the Green's function from secondary point sources distributed on the surface of the rod, we get its impulse response. Then, the displacement field is computed for each angle of rotation of the rod as regard to the fibers from $\theta=0°$ to 180° with a 10° step (Fig. 6). For each angle, the shear wave speed (Fig. 7) is deduced from a spectral analysis of the simulated dis-

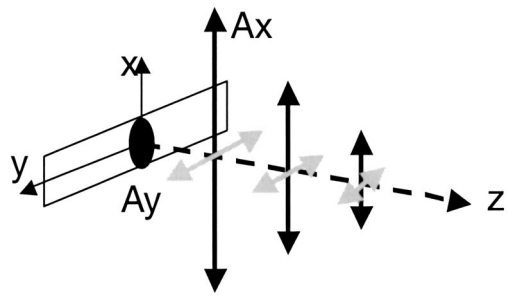


FIG. 5. 3D representation of the amplitude of the displacement gradient. The black and the gray arrows represent the gradient amplitude in the x and y direction, respectively.

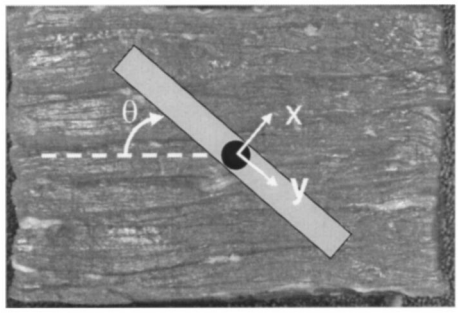


FIG. 6. Top view of an experimental setup. A rod (80 mm long) is used as a shear wave source. The speed is measured along the z axis for angle between the rod and the fibers ranging from $\theta=0°$ to 180°.

placements. The speed parameters used in this numerical simulation are chosen to be as close as possible to the experiment presented in the next section: $V_p=1500\ \mathrm{m\,s^{-1}}$, $V_s^{\perp}=10\ \mathrm{m\,s^{-1}}$, $V_s^{\parallel}=28\ \mathrm{m\,s^{-1}}$. The agreement between these values and the measured speeds is good especially for the two extreme positions (0° and 90°): they are, respectively, $29\ \mathrm{m\,s^{-1}}$ and $10\ \mathrm{m\,s^{-1}}$. The intermediate speeds are a combination of both shear speeds. Consequently these simulation results confirm the assumption that speed measurements with a rod as a shear wave source are very close to the values of polarized plane waves. The crucial point of this paragraph is that the use of a rod favors the strain field in the direction

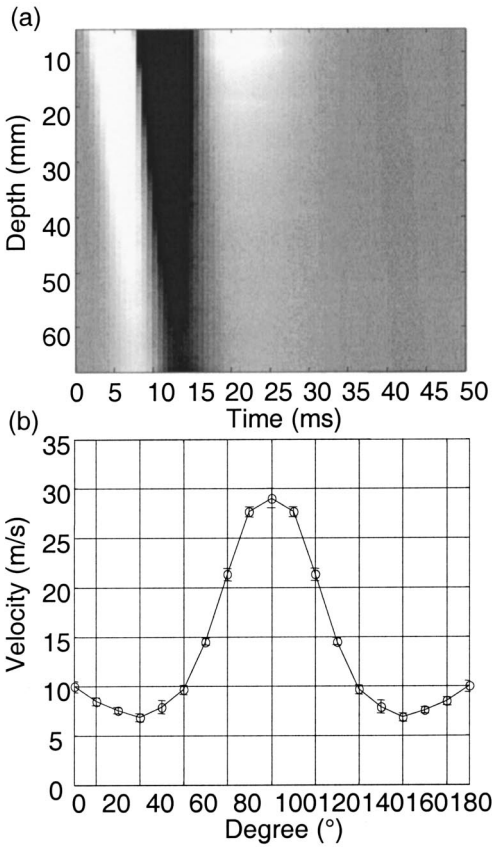


FIG. 7. An acoustic pulse given by a rod on an transverse isotropic media is simulated using the Vavryčuk's theoretical model. (a) Displacement field along depth versus time for $\theta=90°$ between the rod and the muscle fibers model. (b) The shear wave speed as function of the angle of rotation θ.

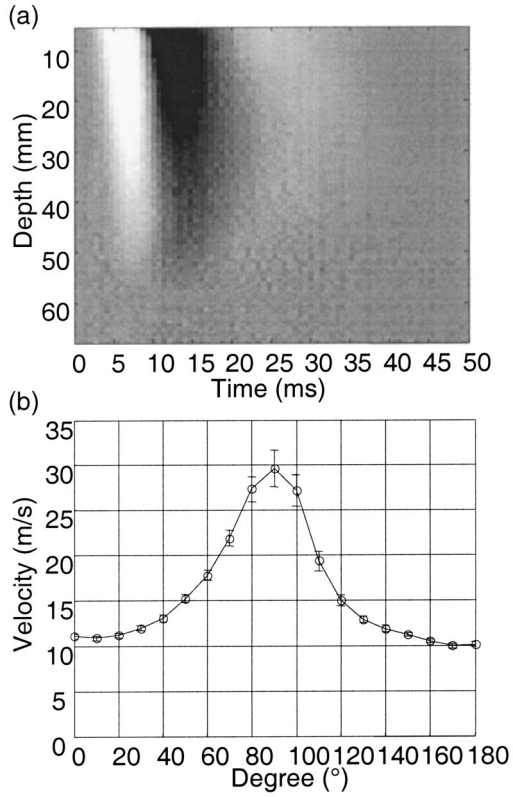


FIG. 8. An acoustic pulse is given by a rod on a beef muscle (*biceps femoris semi-tendinosus*). (a) Experimental displacement field along depth versus time for $\theta = 90°$ between the rod and the muscle fibers. (b) The shear wave speed as function of the angle of rotation θ.

perpendicular [x direction (Fig. 5)]. Nevertheless the polarization of the shear wave in the axis of the source remains longitudinal (z axis).

IV. EXPERIMENTAL RESULTS

A. *In vitro* experiment on beef muscle

In the experimental setup on beef muscle (*biceps femoris semi-tendinosus*), the fibers are clearly visible on the picture (Fig. 6) and well aligned. The ultrasonic equipment is described in the first part. The transducer is set in the middle of the rod (80 mm long) and allows to measure the longitudinal component of the displacement field along the z axis. A low frequency pulse at 100 Hz central frequency is generated with the rod applied on the surface of the sample. The speeds of the shear wave are measured for angles of rotation of the rod with regard to the fibers ranging from $\theta = 0°$ to 180° with a 10° step (as in the numerical experiment). The results (Fig. 8) show that the speed is maximum ($V_s^{\parallel} = 28\ \mathrm{m\,s^{-1}}$) when the rod is perpendicular to the fibers ($\theta = 90°$) and minimum ($V_s^{\perp} = 10\ \mathrm{m\,s^{-1}}$) when the rod is parallel to the fibers ($\theta = 0°$ and $\theta = 180°$). It is in good qualitative agreement with rheology studies on the same kind of muscles[17] since we find c_{66} (110 kPa), the shear elasticity perpendicular to the fibers, to be smaller than c_{44} (862 kPa), the shear elasticity parallel to the fibers. Moreover, the experimental results are in good quantitative agreement with the simulation (Fig. 7) although the viscosity has not been taken into account in our theoretical model.

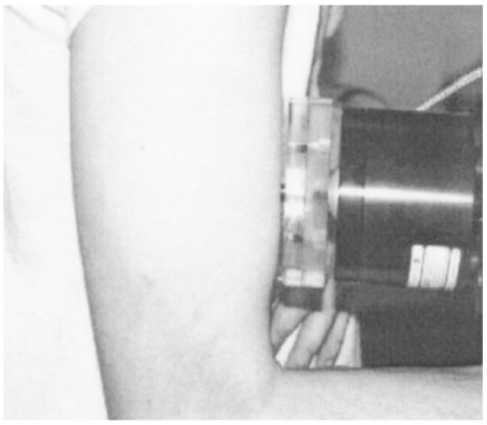


FIG. 9. Experimental setup on male biceps. The probe is applied on the surface of the biceps and the shear wave speed is measured.

B. *In vivo* experiment

The *in vivo* experiment presented here are done on the human biceps. One reason is that this muscle has a simple structure, the fibers are aligned in the same direction. A second reason is that it is often used as a reference in the literature. A volunteer lift at 6 kg charge. On the contracted muscle the rod system (Fig. 9) is applied. As in the previous experiments the shear wave speed is measured for various angles between the rod and the fibers. On the experimental curve of Fig. 10 the speed reaches a maximum when the rod is perpendicular to the fibers ($V_s^{\parallel} = 12\ \mathrm{m\,s^{-1}}$) and a minimum when the rod is parallel to the fibers ($V_s^{\perp} = 3\ \mathrm{m\,s^{-1}}$). These *in vivo* measurements are not surprising given the results of the simulation (Fig. 7) and of the *in vitro* experiments (Fig. 8). They confirm the idea that soft tissues like muscles are very anisotropic. As a comparison, in crystals or rocks, the ratio of the speed of waves in different direction of propagation rarely exceeds 2. It can be noted that the *in vivo* measurements are not as smooth as the *in vitro* experiment. A reason is that one cannot avoid tissue motions during the experiment. Moreover the speed is not measured on a long distance due to the finite size of the biceps (typically 2 cm) which increases the experimental error. Finally, in the case

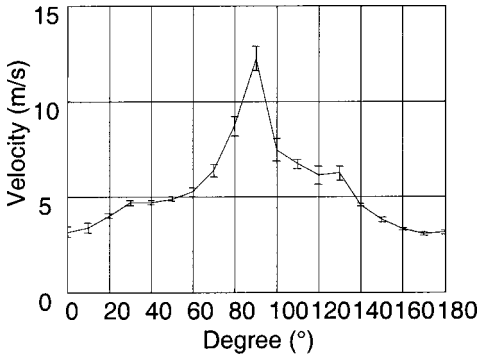


FIG. 10. *In vivo* measurements of the shear wave speed as function of the angle of rotation θ between the rod and the biceps fibers. A ratio of 4 is found between the fast and the slow shear wave which characterizes a strong anisotropy.

where the dimensions of the propagation medium are comparable to the shear wavelength, the boundary conditions may influence the measurements.

V. CONCLUSION

Since anisotropy in soft tissues is a rule more than an exception, elastographic techniques have to take into account this property in the measurement of the elasticity. We have shown in this paper that the use of a low frequency rod source in transient elastography enables one to study a simple transverse isotropic medium: the biceps. The validation of this technique is supported by numerical simulations based on Green's function in isotropic and hexagonal anisotropic solids (transverse isotropy). A strong anisotropy is found on the *in vitro* beef muscle ($V_s^{\perp}=10\ \mathrm{m\,s^{-1}}$, $V_s^{\parallel}=28\ \mathrm{m\,s^{-1}}$) and in the *in vivo* human biceps ($V_s^{\perp}=3\ \mathrm{m\,s^{-1}}$, $V_s^{\parallel}=12\ \mathrm{m\,s^{-1}}$). These results gave birth to a research program with the myology institute of the hospital ''La Pitié Salpetrière'' in Paris.

ACKNOWLEDGMENT

The authors wish to express their grateful thanks to Patrick Rasolofosaon for some illuminating discussions.

APPENDIX: THE GREEN'S FUNCTION IN A HEXAGONAL ANISOTROPIC MEDIUM

In an hexagonal anisotropic, homogeneous and elastic infinite solid, the displacement field of a point source, the Green's function, is derived from higher-order ray theory (Vavryčuk). The solution is given in Cartesian coordinates $\mathbf{x}=(x,y,z)$:

$$G_{kl}(\mathbf{x},t)=\frac{1}{4\pi\rho}\{G_1(\mathbf{x},t)+G_2(\mathbf{x},t)+G_3(\mathbf{x},t)+G_4(\mathbf{x},t)+G_5(\mathbf{x},t)\},\tag{A1}$$

where k and l are the direction index numbers of the displacement and the source, respectively. The other terms are defined as

$$G_1(\mathbf{x},t)\frac{1}{\sqrt{c_{11}^3}}\frac{g_{1k}g_{1l}}{\tau_1}\delta(t-\tau_1),$$

$$G_2(\mathbf{x},t)=\frac{1}{\sqrt{c_{44}^3}}\frac{g_{2k}g_{2l}}{\tau_2}\delta(t-\tau_2),$$

$$G_3(\mathbf{x},t)=\frac{1}{c_{66}\sqrt{c_{44}}}\frac{g_{3k}g_{3l}}{\tau_3}\delta(t-\tau_3),\tag{A2}$$

$$G_4(\mathbf{x},t)=\frac{1}{\sqrt{c_{44}}}\frac{g_{3k}^{\perp}g_{3l}^{\perp}-g_{3k}g_{3l}}{R^2}\int_{\tau_2}^{\tau_3}\delta(t-\tau)d\tau,$$

$$G_5(\mathbf{x},t)=\frac{3g_{1k}g_{1l}-\delta_{kl}}{r^3}\int_{\tau_1}^{\tau_2}r\,\delta(t-\tau)d\tau.$$

Here τ_1, τ_2, τ_3 are the travel time of, respectively, the compressional wave, the slow shear wave and the fast shear wave,

$$\tau_1=\frac{r}{\sqrt{c_{11}}},\tag{A3}$$

$$\tau_2=\frac{r}{\sqrt{c_{44}}},\tag{A4}$$

$$\tau_3=\frac{r}{\sqrt{c_{66}}}\sqrt{N_1^2+\frac{c_{66}}{c_{44}}N_3^2},\quad \tau_3=\frac{r}{\sqrt{c_{66}}}\sqrt{N_1^2+N_2^2+\frac{c_{66}}{c_{44}}N_3^2}.\tag{A5}$$

$R=\sqrt{x^2+y^2}$ is the distance to the receiver from the vertical axis, $r=\sqrt{x^2+y^2+z^2}$ is the distance from the source to the receiver. The polarization vectors are given by the following equations:

$$g_1=[N_1,N_2,N_3],\tag{A6}$$

$$g_2=\frac{-1}{\sqrt{N_1^2+N_2^2}}[-N_1N_3,-N_2N_3,N_1^2+N_2^2],\tag{A7}$$

$$g_3=\frac{1}{\sqrt{N_1^2+N_2^2}}[N_2,-N_1,0],\tag{A8}$$

$$g_3^{\perp}=\frac{1}{\sqrt{N_1^2+N_2^2}}[N_1,N_2,0],\tag{A9}$$

where $N_m=x_m/r$ is the unit direction vector to the receiver.

[1] S. Crampin and S. C. Kirkwood, ''Shear-wave singularities of wave propagation in anisotropic media,'' J. Geophys. **49**, 43–46 (1981).

[2] T. Mensch and P. Rasolofosaon, ''Elastic-wave velocities in anisotropic media of arbitrary symmetry-generalization of Thomsen's parameters ε, δ and γ,'' Geophys. J. Int. **128**, 43–64 (1997).

[3] J. E. Zimmer and J. R. Cost, ''Determination of the elastic constants of an unidirectional fiber composite using ultrasonic velocity measurements,'' J. Acoust. Soc. Am. **47**, 795–803 (1970).

[4] T. Lhermitte and B. Perrin, ''Anisotropy of elastic properties of cross-ply fiber-reinforced composite materials,'' IEEE Ultrasonic Symposium Proc. **2**, 825–830 (1991).

[5] B. K. Hoffmeister, S. M. Handley, S. A. Wickline, and J. G. Miller, ''Ultrasonic determination of the anisotropy of Young's modulus of fixed tendon and fixed myocardium,'' J. Acoust. Soc. Am. **100**, 3933–3940 (1996).

[6] P. L. Kuo, P. C. Li, and M. L. Li, ''Elastic properties of tendon measured by two different approaches,'' Ultrasound Med. Biol. **27**, 1275–1284 (2001).

[7] H. S. Yoon and J. L. Katz, ''Ultrasonic wave propagation in human cortical bone-I. Theoretical considerations for hexagonal symmetry,'' J. Biomech. **9**, 407–412 (1976).

[8] S. F. Levinson, ''Ultrasound propagation in anisotropic soft tissues: The application of linear elastic theory,'' J. Biomech. **20**, 251–260 (1987).

[9] J. Anderson, ''Elasticité musculaire longitudinale et transversale: influence de l'abscence de desmine—Longitudinal and transversal elasticity of muscles: influence of desmin lack,'' thesis (in French), Compiègne University of Technology, France, 2000.

[10] S. F. Levinson, M. Shinagawa, and T. Sato, ''Sonoelastic determination of human skeletal muscle elasticity,'' J. Biomech. **28**, 1145–1154 (1995).

[11] F. Lee, J. P. Bronson, R. M. Lerner, K. J. Parker, S. R. Huang, and D. J. Roach, ''Sonoelasticity imaging: Results in *in vitro* tissue specimens,'' Radiology **181**, 237 (1991).

[12] S. Catheline, F. Wu, and M. Fink, ''A solution to diffraction biases in sonoelasticity: The acoustic impulse technique,'' J. Acoust. Soc. Am. **105**, 2941–2950 (1999).

[13] L. Sandrin, M. Tanter, J. L. Gennisson, S. Catheline, and M. Fink, ''Shear elasticity probe for soft tissue with 1D transient elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 436–446 (2002).

[14] D. C. Gakenheimer and J. Miklowitz, ''Transient excitation of an half space by a point load traveling on the surface,'' J. Appl. Mech. **36**, 505–514 (1969).

[15] V. Vavryčuk, ''Exact elastodynamic Green's functions for simple types of anisotropy derived from higher-order ray theory,'' Stud. Geophys. Geod. **45**, 67–84 (2001).
[16] L. Sandrin, ''Elastographie impulsionnelle par ultrasons: du palpeur acoustique à l'imagerie ultrarapide—Ultrasound based transient elastography: from the shear elasticity probe to the ultrafast imaging,'' thesis (in French), Paris VI University, France, 2000.
[17] J. Lepetit, ''Deformation of collagenous, elastin and muscle fibres in raw meat in relation to anisotropy and length ratio,'' Meat Sci. **26**, 47–66 (1989).

Applications of laser-based ultrasonics to the characterization of the internal structure of teeth

D. W. Blodgett[a)]
The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723

(Received 17 January 2002; revised 20 March 2003; accepted 7 April 2003)

Dental health care and research workers require a means of imaging the structures within teeth *in vivo*. For example, there is a need to image the margins of a restoration for the detection of poor bonding or voids between the restorative material and the dentin. In addition, a high-resolution imaging modality is needed to detect tooth decay in its early stages. If decay can be detected early enough, the process can be monitored and interventional procedures, such as fluoride washes and controlled diet, can be initiated to help remineralize the tooth. Currently employed x-ray imaging is limited in its ability to visualize interfaces and incapable of detecting decay at a stage early enough to avoid invasive cavity preparation followed by a restoration. To this end, nondestructive and noncontact *in vitro* measurements on ''as-is'' extracted sections of human incisors and molars using laser-based ultrasonics are presented. Broadband ultrasonic waves are excited in the extracted sections by using a pulsed carbon-dioxide (CO_2) laser operating in a region of high optical absorption in the dental hard tissues. Optical interferometric detection of the ultrasonic wave surface displacements is accomplished with a path-stabilized Michelson-type interferometer. Laser ultrasonics is found effective in characterizing the anisotropic and inhomogeneous nature of dentin. In addition, time-of-flight analysis of the measured bulk transmission waveforms allows for detection of dentino-enamel and carious dentin-dentin junctions. These results are compared to those obtained for specially prepared tooth phantoms that mimic the mechanical properties of dental hard tissues. [DOI: 10.1121/1.1578080]

PACS numbers: 43.80.Qf, 43.80.Vj, 43.80.Jz [FD]

I. INTRODUCTION

Attempts to apply conventional ultrasonic techniques to the examination of the internal structure of a tooth have met with mixed success.[1–4] One of the major obstacles is finding a suitable couplant for the transducer to the tooth for *in vivo* measurements. Without proper coupling, it is difficult to transfer acoustic energy into the tooth. Early investigators[2,3] attempted to use water, as with soft tissue structures, but the results were not convincing. Reich *et al.*[4] overcame this problem by replacing water with mercury. Although providing superior coupling efficiency, it would never be suitable for clinical applications. Barber *et al.*[1] employed a small aluminum buffer rod to transfer the acoustic energy from the contact transducer to the tooth. They estimated a transmission efficiency of almost 87% using this technique, compared to only 5% using water. Although eliminating the coupling problem, the buffer rod required that a small flat spot be ground in the surface of the tooth, making *in vivo* measurements impossible and limiting spatial resolution. The use of a specially designed transducer that generated a single pulse with a 10-ns rise time allowed the researchers to clearly resolve the dentino–enamel and dentino–pulp interfaces (junctions) of human and bovine incisors. To avoid special surface preparation, Yanikoglu *et al.*[5] used a conventional longitudinal 1.5-mm-diam contact transducer with glycerine as a couplant to image white spot caries. Ghorayeb and Valle[6] avoided special surface preparation by making A- and B-scans on extracted human teeth submerged in water. This further verifies the sensitivity of ultrasonics to dental hard tissue interfaces, but does not lend itself to *in vivo* applications.

Ng *et al.*[7] used an acoustic microscope to improve the spatial resolution of measurements on unblemished and demineralized enamel from extracted human teeth. As with previous researchers, they used water as a couplant, which limited them to inspection depths of about 0.5 to 1.5 mm. Peck *et al.*[8] also took advantage of the increased spatial resolution of the acoustic microscope to detect small caries lesions in sections of human enamel. In addition, they were able to quantify the anisotropic nature of the enamel via variations in the Raleigh velocity. However, as with the previous work by Barber *et al.*,[1] special polishing of the tooth samples was required, again making the technique ill-suited for clinical applications.

In this work, laser-based ultrasonics is investigated for characterizing the internal structure of dental hard tissues. Laser ultrasonics overcomes previous difficulties in coupling efficiency, need for special surface preparation, and detection footprint size by using a short-pulse laser, in place of a contact transducer, to generate high-frequency (broad-band) ultrasound in a material.[9,10] Optical interferometric detection of the ultrasound by a Michelson-type path-stabilized interferometer provides a complementary technique for noncontact sensing of ultrasonic waves. Specifically, *in vitro* measurements are presented for sections of extracted human molars. These results are compared with those obtained on specially prepared tooth phantoms with mechanical interface

a)Electronic mail: david.blodgett@jhuapl.edu

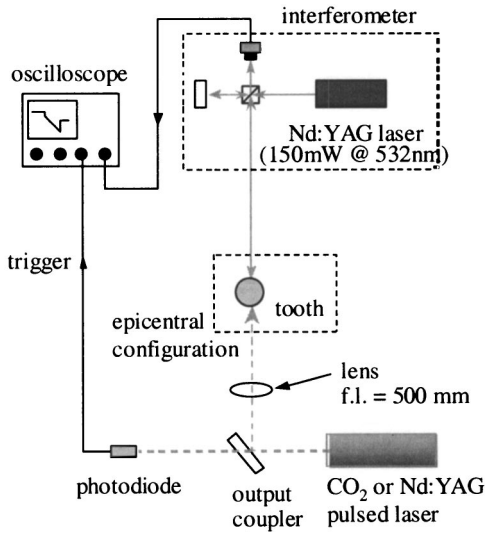


FIG. 1. Experimental setup for laser ultrasonic system.

properties similar to those found in teeth. In addition, results from extracted human teeth are presented to show that laser-generated ultrasound in sensitive to the anisotropic and inhomogeneous nature of dentin both within a single tooth and between different teeth. Finally, ultrasonic measurements are presented on an extracted human molar with surface caries located between the cementum/enamel junction.

II. METHODS AND MATERIALS

A. Experimental configuration

Ultrasound measurements were taken in an epicentral configuration (source and receiver on opposite sides—see Fig. 1). A pulsed CO_2 laser with a 100-ns pulse at full-width half-maximum (FWHM) at 10.6 μm (Edinburgh Instruments Ltd. μTEA laser) was used to generate the ultrasonic waves in the tooth sections. For the tooth phantoms, a pulsed Nd:YAG laser with an 18-ns pulse at FWHM and 1.064 μm (Spectra Physics) was used for ultrasound generation due to the poor absorption properties of aluminum in the tooth phantoms at 10.6 μm. In both cases, the pulsed laser was focused onto the back of the specimen using a 500-mm focal length lens. Assuming near diffraction-limited performance, the focused spot size radius, w_{diff}, can be estimated at 1 mm for the CO_2 laser and 100 μm for the Nd:YAG laser. The output power of the pulsed lasers was controlled using polarization optics to ensure nondestructive (thermoelastic) acoustic wave generation.

A path-stabilized Michelson-type interferometer with a 150-mW diode-pumped frequency-doubled Nd:YAG laser source at 532 nm (Coherent DPSS 532) was used to detect the ultrasonic wave arrivals in both the tooth sections and tooth phantoms.[9,10] To account for the small amount of reflected light from the front surface of the tooth, an interferometer design was chosen so that the division of optical power between the reference and object arms could be controlled through use of a polarization-based scheme. Using this configuration, the optical power reflected by the specimen was increased enough to provide an adequate signal level. Due to the nonspecular nature of the tooth surface, a microscope objective (20×, 0.40 N.A.) was used to both focus and collect the reflected light. A microscope objective allowed the interferometer laser to be focused to a spot size approaching 3 μm. The output of the interferometer was passed to a high-speed digitizing oscilloscope (Lecroy 9384) for recording the ultrasonic event. The inherently flat frequency response of the Michelson interferometer allowed for actual surface displacement amplitudes to be determined.[9] Capture of the ultrasonic waveform was triggered by a high-speed photodetector that monitored a sampling of the output pulse of the pulsed laser. Sampling of the output laser pulse allowed for very accurate time-of-flight measurements for the acoustic waves. Multiple waveforms were averaged 100 times to improve the signal-to-noise ratio (SNR).

B. Acoustic wave generation

Laser-generated ultrasound uses a short-pulse laser, in place of a contact transducer, to generate high-frequency (broad-band) ultrasound in a material.[9,10] Due to the absorption of pulse energy at or near the surface of the specimen, temperature gradients are established within the material, producing a rapidly changing strain field. This strain field, in turn, radiates energy as elastic (ultrasonic) waves. At low pulse energies, this is an entirely thermoelastic process, resulting in no damage to the material under test. Efficient ultrasound generation depends upon the material's absorption characteristics at the optical wavelength of the pulsed laser.[9] Work by Zuerlein *et al.*[11] and Duplain *et al.*[12] have shown that enamel and dentin have strong absorption in the 9–11-μm region due to the phosphate in the carbonated hydroxyapatite (CAP). Zuerlein *et al.*[11] determined absorption coefficients of 5500, 8000, 1125, and 825 cm^{-1} at 9.3, 9.6, 10.3, and 10.6 μm, respectively. These correspond to absorption depths between 1.25 and 12 μm.

For laser generation to be acceptable for future *in vivo* applications, the total temperature rise at the tooth surface must be small enough to ensure that no damage is incurred. Melting of the individual enamel prisms is known to occur between 800 °C and 1200 °C and is inversely related to the carbonate content.[11] The presence of C cracks in enamel has been found at temperatures as low as 150 °C when exposed for periods approaching 1 h.[13] The actual temperature rise in the enamel can be estimated for one-dimensional, thermoelastic ultrasonic wave generation if it is assumed that the optical pulse is so short that the thermal conductivity into the bulk of the enamel can be neglected, and the absorption takes place uniformly down to a minimum depth equal to the skin depth.[9] For a pulse with a Gaussian temporal profile, the temperature rise, T, as a function of depth, z, and time, t, is given by

$$T(z,t)=\frac{\sqrt{\kappa}}{K\sqrt{\pi}}\int_0^t \frac{(1-R)I_0 e^{\xi(t-t')^2/\tau^2} e^{-z^2/4\kappa t'}}{\sqrt{t'}}\,dt'. \quad (1)$$

In this expression, K is the thermal conductivity, κ is the thermal diffusivity, ξ is a fitting parameter ($\xi=2.7$) for the pulse temporal width, τ is the pulse width, R is the measured reflectivity at the wavelength ($R=13.8\%$[11]), and I_0 is the maximum power density. The thermal properties of the dif-

TABLE I. Acoustic and thermal properties for the different layers in a tooth as well as acoustically similar materials.[10,24]

Layer	V_{long} (mm/μs)	V_{shear} (mm/μs)	Acoustic impedance Z [$\times 10^6$ kg/m^{2}s]	Thermal cond. (W/cmK)	Thermal diff. (mm^2/s)
Enamel	6.25	3.10	18.8	9.34×10^{-3}	0.469
Aluminum	6.30	3.10	17.0	2.37	97.8
Dentin	3.80	1.90	7.6	5.82×10^{-3}	0.183
Copper	4.70	2.30	41.6	4.01	116.2
Amalgam	4.35	2.26	33.7	0.226	
Borosilicate glass	5.30	3.00	18.9	0.014	0.795

ferent thermal hard tissues are listed in Table I. In this work, the pulse power is about 8 mJ and the laser focused to a 1-mm-radius spot size corresponding to a maximum power density of 2.2×10^{10} Wm^{-2}. Figure 2 shows a plot of the temperature evolution at depths of 5, 20, and 100 μm below the surface. This analysis shows that the expected temperature rise 5 μm below the surface approaches 100 °C after 30 μs and only exceeds 60 °C for about 150 μs. At a depth of 20 μm, the temperature never exceeds 25 °C. These temperatures are below those necessary to damage to the enamel structure. This conclusion is further validated by the lack of visible damage to the tooth surface after making ultrasonic measurements. Another consideration is the maximum temperature of the pulp cavity, since studies have shown that the pulp temperature rise should never exceed 5 °C.[14] Calculations show that at a depth of 100 μm the enamel temperature never increase more than 5 °C and that at a depth of 1 mm (a typical enamel thickness), the peak temperature rise do not exceed 0.05 °C.

C. Acoustic wave propagation

The anisotropic nature of dental hard tissue is well-documented.[6,15,16] Peck *et al.* used an acoustic microscope to quantify variations in the elastic properties of enamel by measuring the Rayleigh velocity as a function of enamel rod orientation.[6] They found that the Rayleigh velocity varied by only 5% in these measurements. Although dentin is the hard tissue that constitutes the body of the tooth, its mechanical properties are not well-understood.[17,18] However, the presence of dentinal tubules would be expected to cause some degree of directional anisotropy. For each of these dental hard tissues, inhomogeneity is also expected to play a role in variations both within a tooth and between different teeth.

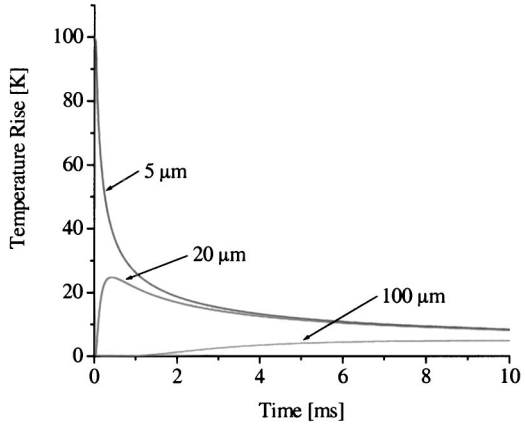


FIG. 2. Calculated temperature distribution in enamel for a Gaussian pulse with a pulse width of 100 ns and a maximum power density of 2.2×10^{10} Wm^{-2}.

Due to the relatively small amount of anisotropy in the dental hard tissues, we will proceed as with previous researchers[1–5,7,8] and assume that the enamel and dentin are elastically isotropic, and that the analysis of the results can be based on one-dimensional, time-of-flight calculations. For isotropic materials, only two bulk waves must be considered (longitudinal and shear), each with their own characteristic wave speed (ν_{long} and ν_{shear}). When assessing any multiphase structure, such as a tooth, both the acoustic velocity and acoustic impedance, Z, of each layer must be considered. The intensity of the reflected acoustic radiation, I_r, at the interface between two different media with acoustic impedances of Z_1 and Z_2, is given by

$$I_r=\left(\frac{Z_2-Z_1}{Z_2+Z_1}\right)^2 I_i=\Gamma I_i, \tag{2}$$

where Γ is the reflection coefficient and I_i is the intensity of the incident radiation. Knowledge of the relative reflected and transmitted intensities at each material interface in a tooth structure aids in the interpretation of the final ultrasonic waveform. The acoustic velocities and densities for different components of a tooth are listed in Table I.

Acoustic scattering is an issue for any of the dental hard tissues. Laser generation provides broadband ultrasonic waves, whose peak frequency can be estimated as the reciprocal of the temporal pulse width. For the CO_2 laser, with a 100 ns pulse, the maximum acoustic frequency is about 10 MHz. For enamel and dentin this corresponds to longitudinal acoustic wavelengths of 625 and 380 μm, respectively. At 1 MHz, the wavelength increases to 6.25 and 3.8 mm. Enamel is composed of rods that extend from the dentin-enamel junction (DEJ) to the tooth surface. These rods have a maximum width of about 5 μm and a length approaching 9 μm. Dentin tubules range from about 1 μm in diameter at the outer dentin region to between 3 and 4 μm at the pulpal border. The density of the tubules varies with location in the tooth, with typical separations on the order of a few micrometers. These dimensions indicate that Rayleigh scatter

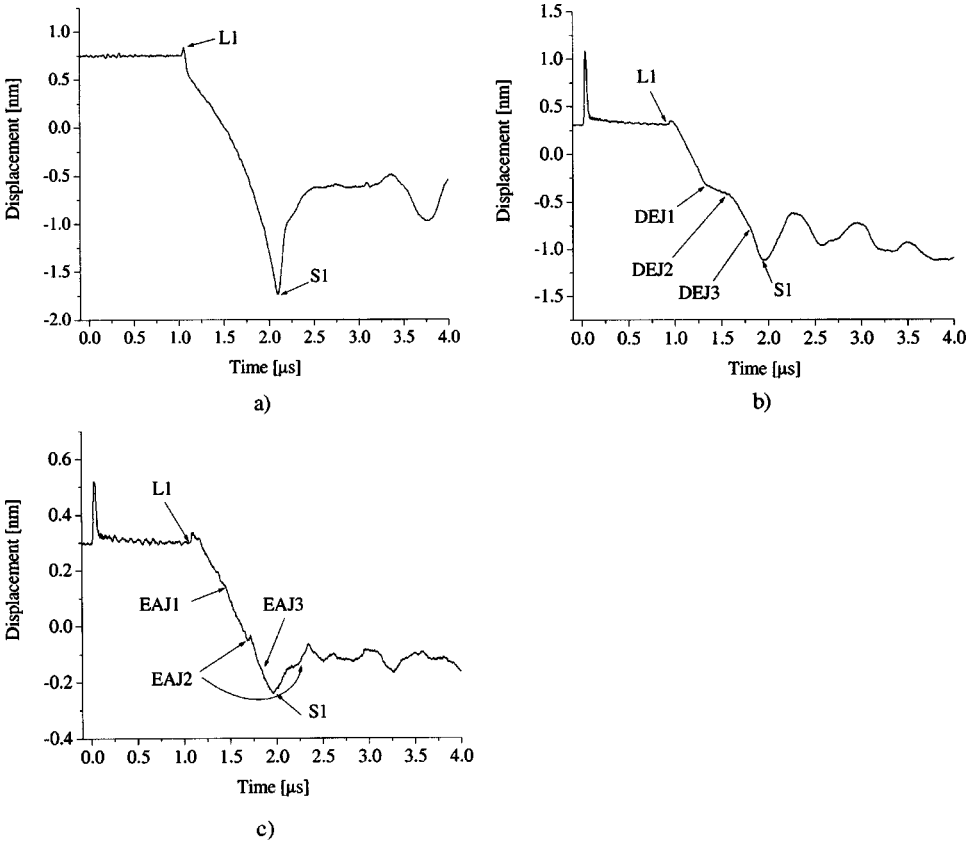


FIG. 3. Comparison of ultrasonic waveforms taken in tooth phantoms mimicking (a) enamel, (b) enamel/dentin/enamel, and (c) enamel/amalgam/enamel.

should play a role at the highest frequencies, and that the effect will be less noticeable at lower frequencies.

D. Tooth section and phantom preparation

One means of better understanding the acoustic signature obtained from an actual tooth structure is to construct a tooth phantom made from acoustically similar materials. To simplify construction, a separate tooth phantom was designed for each of three different acoustic ''paths'' through the tooth section. These paths include enamel, enamel/dentin/enamel, and enamel/amalgam/enamel. Materials with similar acoustic velocities and impedances to the different components of a real tooth are listed in Table I. Since the purpose of this study was to better understand the interface properties between the different dental tissues, similarities in acoustic impedance were viewed as more important than acoustic velocities, so that reflections at interfaces could be more accurately simulated. The reflection coefficient at the enamel/dentin interface is 0.18 vs 0.176 for the tooth phantom and at the enamel/amalgam interface is 0.081 versus 0.026. These reflection coefficient calculations do not account for the effects of the low Z adhesive used to bond the different layers. To decrease the effects of the low-Z interface, the phantom materials were machined flat, and then clamped to minimize the thickness of the interface. Previous work by Lees[19] has shown that low-Z interfaces less than 5 μm thick had minimal effect on the transmitted ultrasonic wave.

Laser-based ultrasonic measurements were also performed on extracted sections of human incisors and molars. Prior to the measurements, the tooth sections were stored in a physiological saline to help preserve the mechanical properties of the dental hard tissues. The first sets of measurements were conducted on sections of dentin. For these measurements the occlusal surface enamel and pulpal region were removed with a wet diamond saw, leaving sections between 2 and 3 mm thick. This configuration allowed for variations in the acoustic wave speeds in dentin to be determined. The second sets of measurements that are presented were performed on a molar that was sectioned vertically through the apex of the tooth so that variations across the dentin/enamel interface could be characterized. The final measurements were taken on an extracted molar with visible caries between the cementum/enamel junction. As with the previous measurements, the tooth was sectioned vertically through the apex of the tooth. The purpose of these measurements was to determine the sensitivity of ultrasonics to the carious dentin/dentin interface.

III. RESULTS

A. Tooth phantom results

Shown in Fig. 3(a) is a measured thermoelastic ultrasonic waveform from a piece of aluminum (6.45 mm thick). This represents an ideal waveform through either enamel or dentin, assuming isotropic behavior. In this figure, the first longitudinal (L1) and shear wave (S1) arrivals at 1.02 and 2.08 μs are marked. The aluminum waveform also provides a baseline for the other two tooth phantom waveforms [Figs. 3(b) and (c)]. Each of these waveforms is distinctly differ-

ent due to reflections at the interfaces of the different layers of the tooth phantoms. Each waveform was averaged 100 times to improve the SNR.

The measured waveform from the second tooth phantom [Fig. 3(b)] simulates results from the dentin/enamel junction (DEJ). This phantom is composed of a 1.95-mm-thick piece of copper (dentin) sandwiched between two pieces of aluminum (enamel) that are 1.95 and 1.25 mm thick. The first longitudinal arrival time, t_{L1}, occurs at

$$t_{L1} = \frac{d1_{\text{aluminum}}}{\nu_{\text{aluminum}}} + \frac{d_{\text{copper}}}{\nu_{\text{copper}}} + \frac{d2_{\text{aluminum}}}{\nu_{\text{aluminum}}}$$
$$= \frac{1.95\ \text{mm}}{6.3\ \text{mm}/\mu\text{s}} + \frac{1.95\ \text{mm}}{4.7\ \text{mm}/\mu\text{s}} + \frac{1.25\ \text{mm}}{6.3\ \text{mm}/\mu\text{s}}$$
$$= 0.923\ \mu\text{s}. \quad (3)$$

The second acoustic wave arrival occurs when the longitudinal wave traverses the thin piece of aluminum (enamel) a second time after reflection at the DEJ. This second arrival, t_{DEJ1}, occurs at

$$t_{DEJ1} = t_{L1} + \frac{d2_{\text{aluminum}}}{\nu_{\text{aluminum}}}$$
$$= 0.923\ \mu\text{s} + 0.397\ \mu\text{s} = 1.32\ \mu\text{s}. \quad (4)$$

The next two acoustic wave arrivals, DEJ2 and DEJ3, result from additional traversals of the thicker piece of aluminum and the copper sections.

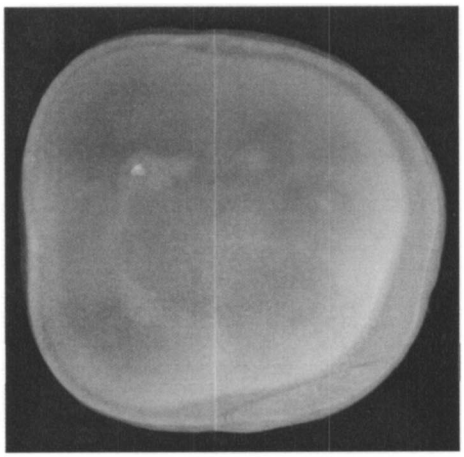
The ultrasonic waveform for the aluminum/glass/aluminum (enamel/amalgam/enamel) tooth phantom is shown in Fig. 3(c). The first longitudinal arrival (L1) is identified by the initiation of the positive slope in the waveform. As with the previous phantom, this arrival corresponds to a direct acoustic path for the longitudinal wave through the entire tooth phantom. In this tooth phantom, the first section of aluminum (enamel) is 1.95 mm thick, the glass (amalgam) is 3.3 mm thick, and the final section of aluminum (enamel) is 1.25 mm thick. This leads to an expected arrival time of 1.13 μs. The second longitudinal arrival (EAJ1) occurs after the longitudinal wave traverses the thinner section of enamel (aluminum) a second time after reflection at the enamel-amalgam (aluminum/glass) junction (EAJ) 0.4 μs after L1. Subsequent longitudinal wave arrivals are also visible due to additional traversals of the thicker piece of enamel (aluminum) and amalgam (glass) after reflections at the EAJ. The amplitude of each of these acoustic wave arrivals is smaller than in the previous tooth phantom due to the closer acoustic impedance match of the materials and consequent smaller reflection coefficient ($\Gamma = 0.026$). The EAJ would be more visible in an actual tooth due to the larger reflection coefficient ($\Gamma = 0.081$).

B. *In vitro* human tooth section results

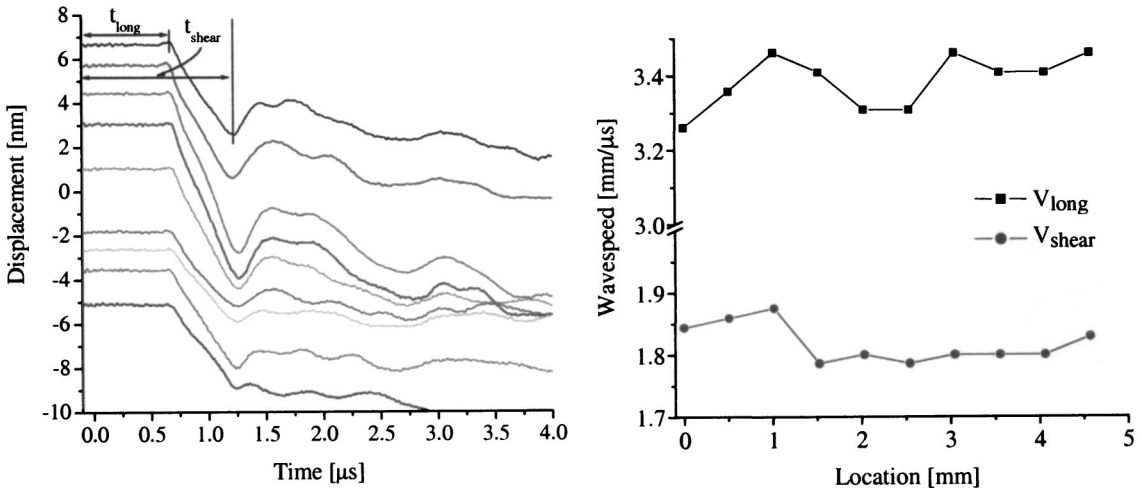
Figure 4(a) shows a picture of one of the extracted tooth sections (2.25 mm thick) with the enamel ring visible around the dentin. Measurements were taken across the middle of the section at 500-μm intervals. Shown in Fig. 4(b) are the measured ultrasonic waveforms taken from the left (top waveform) to right (bottom waveform) of the section. The first measurement was taken about 2 mm to the right of the enamel to avoid boundary effects. The longitudinal and shear wave arrivals are marked in the top waveform. These waveforms are very similar to those found for the elastically isotropic aluminum sample shown in Fig. 3(a). The corresponding calculated wave speeds are shown in Fig. 4(c) and vary less than 5% across the tooth section. The dentinal tubules range from being parallel to subtending an angle of less than 45° with respect to the wave propagation direction. Over 100 total measurements were made from ten different sections, and the average longitudinal and shear wave speeds were 3.35 and 1.83 mm/μs, respectively. The total variation in acoustic wave speeds was found to be less than 10% over all of these measurements. The 10% variations are similar to those found by Lertchirakarn *et al.* in their measurements of the ultimate tensile strength of dentin parallel to the dentinal tubules.[20] These longitudinal wave speeds are slightly lower than the 3.8 ± 0.37 mm/μs reported by Barber *et al.*[1]

Additional measurements on an extracted molar sectioned through the apex of the tooth [see Fig. 5(a)] are presented. This section allowed for the interface properties at the dentin/enamel junctions to be characterized. For these measurements, the ultrasound was generated on the sectioned dentin side and the interferometric measurements on the ''as-is'' enamel surfaces. The corresponding measured ultrasonic waveforms are shown in Fig. 5(b). These measurements were taken at 0.5-mm increments, starting about 1 mm below the enamel on the top portion of the tooth. In this region, the enamel and dentin are about 0.84 and 2.54 mm thick, respectively. The first longitudinal arrival (L1) varies between 0.83 and 0.9 μs for the different waveforms. The first reflection from the DEJ occurs between about 1 and 1.09 μs. These variations are attributed to differences in both acoustic wave speeds and dental hard tissue thicknesses. Using the longitudinal wave speed values for enamel and dentin reported in Table I, the expected arrival times are about 0.8 and 1.08 μs. In the fourth waveform, there are two visible DEJ reflections (DEJ1 and DEJ2). The second DEJ reflection at 1.4 μs is not clearly visible in the other waveforms. The expected arrival time for the second reflection is 1.35 μs.

The final sets of measurements were conducted on an extracted molar with caries on the dentin between the cementum/enamel junction on the surface of the tooth [see Fig. 6(a)]. For these measurements, the ultrasound was generated on the exterior dentin. As with Ghorayeb and Valle[8] and F. C. Yanikoglu *et al.*,[21] transmission measurements were made through the caries to characterize the interface properties between healthy dentin and carious dentin. Also visible in the figure is an open pulp cavity. Shown in Fig. 6(b) are measured ultrasonic waveforms through the caries region (top), through healthy dentin below the pulp cavity (middle), and just above the pulp cavity (bottom). For the caries section, four reflections from the caries/dentin junction (CDJ) occur after the first longitudinal wave arrival (L1). The successive temporal spacing between each of these arrivals is about 0.3 μs. The loss of clarity in the later reflections is attributed to attenuation of the higher frequency acoustic waves during multiple passes through the caries re-

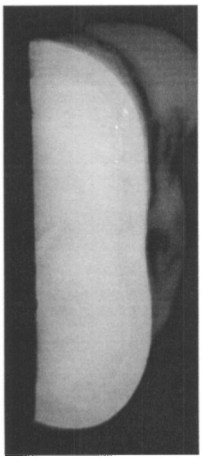


a)

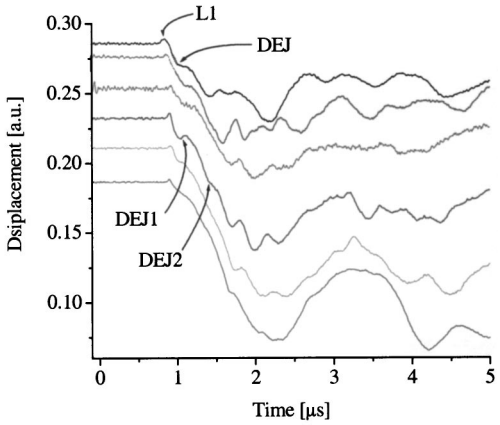


b) c)

FIG. 4. Variations in the acoustic wave speeds in dentin due to anisotropy and inhomogeneity were measured in specially prepared dentin sections. Shown in (a) is a picture of one of the dentin sections with the location of the ultrasonic measurements marked. Measured ultrasonic waveforms across the dentin section taken at 500 μm intervals are shown in (b) with a plot of the calculated wave speeds shown in (c).



a)



b)

FIG. 5. Variations in ultrasonic waveforms were measured in a section of extracted molar. Shown in (a) is an image of the sectioned molar through which the ultrasonic measurements were made showing the DEJ. Selected waveforms measured through the DEJ are shown in (b).

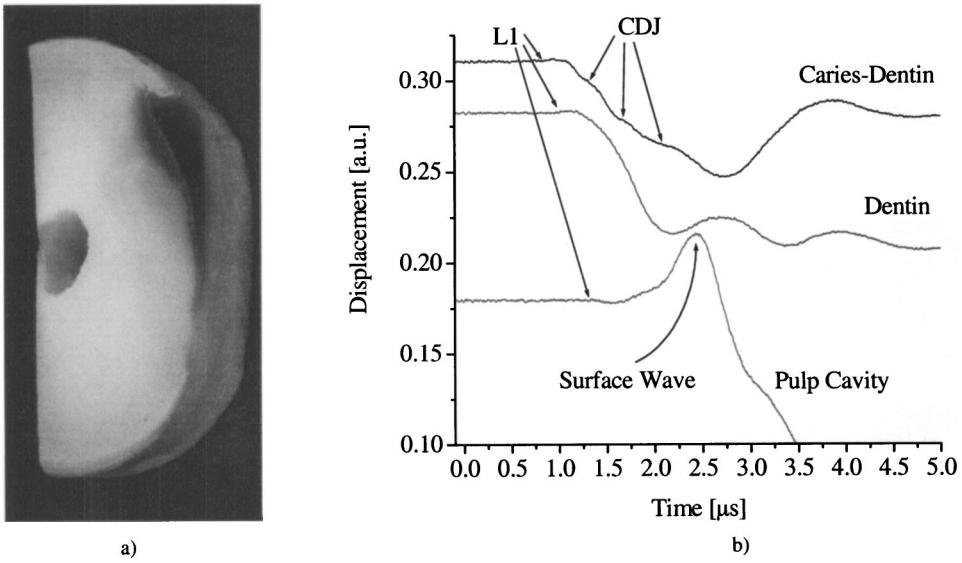


FIG. 6. The ability to resolve depth of carious regions in teeth was tested by making measurements across an exposed cavity between the cemento-enamel junction. Shown in (a) is a picture of the exposed cavity with the measured waveforms through the carious region, healthy dentin, and beside the pulp cavity shown in (b).

gion. This is not unexpected since the presence of caries results in a porosity increase in the dental hard tissue. The second waveform is taken through the relatively healthy dentin to the right of the pulp cavity. This waveform is similar to those shown in Fig. 4(b) with distinct longitudinal (L1) and shear (S1) wave arrivals. The final waveform was measured just to the left of the extirpated pulp cavity. This waveform is distinctly different from the other waveforms, with the pronounced surface wave arrival at 2.43 μs. The presence of the surface wave is attributed to mode conversions at the dentin–pulp cavity interface. The displacement amplitude of the surface wave is substantially larger than either longitudinal or shear acoustic wave arrivals.

C. Error analysis

Aussel and Monchalin[22] estimated the error in making laser ultrasonic measurements at less than 1% when considering phenomena such as acoustic diffraction and timing precision from laser alignment. For measurements on specially prepared samples, they estimated an error of 0.08% in thickness measurements. These errors are representative of what would be found for the tooth phantom measurements.

Measurements on the extracted tooth sections are less precise due to variations in the wave speeds in the dental hard tissues. If one assumes that the uncertainties in the wave speeds in enamel and dentin are 5% and 10%, respectively, and that any uncertainties in timing and alignment can be neglected, the error in thickness determination, δd, can be estimated with[23]

$$d+\delta d=(\nu_{\text{enamel}}t_{\text{enamel}}+\nu_{\text{dentin}}t_{\text{dentin}}) +(\delta\nu_{\text{enamel}}t_{\text{enamel}}+\delta\nu_{\text{dentin}}t_{\text{dentin}}). \quad (5)$$

For a 1-mm section of enamel and a 3-mm section of dentin, there would be an uncertainty in total thickness of about 0.3 mm.

Even with the uncertainties in acoustic wave speeds, the measurements presented here show that dentin/enamel and carious dentin/non-carious-dentin junctions are discernable, and that estimates of the thickness of each of these structures can be made. In addition, uncertainties in dental structure thickness do not affect the ability to detect any cracks or voids within the tooth. Cracks in dental hard tissues or voids between dental hard tissues and a restoration material will be much easier to detect than junctions between dental hard tissues due to the large acoustic impedance mismatch. For these cases, it is the presence of the crack or void that is important, not the exact location.

IV. SUMMARY

The application of laser ultrasonics to the *in vitro* assessment of the internal structure of teeth has been presented. In this technique, a short-pulse CO_2 laser operating in a region of high optical absorption in dental hard tissues is used to nondestructively generate broadband, high-frequency acoustic waves. As the acoustic waves propagate, they are reflected/transmitted at the various interfaces within the tooth structure. These acoustic waves are then detected using an optical interferometric technique. Knowledge of the various acoustic wave velocities and reflection coefficients allow for the internal structure of the tooth structure to be inferred. As with previous researchers, ultrasound has been shown to be sensitive to variations in the acoustic wave speeds in dental hard tissues and to the presence of dentin/enamel junctions and carious/non-carious dentin. However, this is the first known application of laser-based ultrasonics to the *in vitro* characterization of dental hard tissues. The disadvantage of previous techniques is that they either require special surface preparation or lack spatial resolution. Laser ultrasonics does not require any special surface preparation as demonstrated by the ability to make measurements on either the exterior or sectioned portion of a tooth. In addition, the relatively low temperatures generated by the pulsed laser in the dental enamel appear to be safe for both the dental hard and soft tissues. When considering spatial resolution, it is the size of the interferometric spot that is critical. Measurements in this work use spot sizes approaching 3 μm. Spatial resolution is critical when trying to resolve small subsurface defects on the order of 0.1 mm, such as an incipient lesion. As men-

tioned above, another important application is the detection of cracks in dental hard tissues or voids between dental hard tissues and an amalgam restoration.

The *in vitro* measurements presented in this work were made on sections of extracted human incisors and molars. This configuration simplified waveform interpretation since the acoustic wave propagated through a maximum of one interface. In addition, the interface was nearly parallel to the direction to the wave propagation direction, which led to minimal mode conversion. Both mode conversions and acoustic attenuation from the dental structure would make interpretation of complete tooth measurements more difficult. To simplify the analysis, one would switch to same-side acoustic wave generation and detection and look at reflections from interfaces instead of transmission through interfaces. This is what is currently being done by researchers using contact and immersion transducers. It is believed laser-based ultrasonics, when operated in the same-side configuration, has the potential to detect early tooth decay, cracks in dental hard tissues, and voids between dental hard tissues and restorations.

ACKNOWLEDGMENTS

This research was funded under a Johns Hopkins University/Applied Physics Laboratory IR&D program. The author would also like to acknowledge Dr. Ward Massey at the Baltimore College of Dental Surgery, Dental School and Dr. Christopher Longbottom at the University of Dundee for many helpful discussions and for providing the extracted teeth used for this study.

[1] F. E. Barber, S. Lees, and R. R. Lobene, ''Ultrasonic Pulse-Echo Measurements in Teeth,'' Arch. Oral Biol. **14**, 745–760 (1969).
[2] G. Baum, I. Greenwood, S. Slawski, and R. Smirnov, ''Observation of internal structures of teeth by ultrasonography,'' Science **139**, 495–496 (1963).
[3] G. Kossoff and C. J. Sharpe, ''Examination of the contents of the pulp cavity in teeth,'' Ultrasonics **4**, 77–83 (1966).
[4] F. R. Reich, B. B. Brenden, and N. S. Porter, ''Ultrasonic imaging of teeth,'' *Report of Battelle Memorial Institute*, Pacific Northwest Laboratory (1967).
[5] F. C. Yanikoglu, F. Ozturk, O. Hayran, M. Analoui, and G. K. Stookey, ''Detection of natural white spot caries lesions by an ultrasonic system,'' Caries Res. **34**, 225–232 (2000).
[6] S. R. Ghorayeb and T. Valle, ''Experimental evaluation of human teeth using noninvasive ultrasound: echodentography,'' IEEE Trans. Ultrason. **49**(10), 1437–1443 (2002).
[7] S. Y. Ng, M. W. J. Ferguson, P. A. Payne, and P. Slater, ''Ultrasonic studies of unblemished and artificially demineralized enamel in extracted human teeth: a new method for detecting early caries,'' J. Dent. **18**, 201–209 (1987).
[8] S. D. Peck, J. M. Rowe, and G. A. D. Briggs, ''Studies on Sound and Carious Enamel with the Quantitative Acoustic Microscope,'' J. Dent. Res. **68**, 107–112 (1998).
[9] C. Scruby and L. Drain, *Laser Ultrasonics: Techniques & Applications* (Hilger, New York, 1990).
[10] J. W. Wagner, ''Optical Detection of Ultrasound,'' in *Physical Acoustics*, edited by W. P. Mason and A. D. Pierce (Academic, New York, 1990), Vol. 19, pp. 201–266.
[11] M. Zuerlein, D. Fried, J. Featherstone, and W. Seka, ''Optical properties of dental enamel in the mid-IR determined by pulsed photothermal radiometry,'' IEEE J. Sel. Top. Quantum Electron. **5**(4), 1083–1089 (1999).
[12] G. Duplain, R. Boulay, and P. A. Belanger, ''Complex index of refraction of dental enamel at CO_2 laser wavelengths,'' Appl. Opt. **26**(20), 4447–4451 (1987).
[13] M. Muller, M. F. Berytrand, G. Quatrehomme, M. Bolla, and J. P. Rocca, ''Macroscopic and microscopic aspects of incinerated teeth,'' J. Forensic Odontostomatol **16**(1), 1–7 (1998).
[14] L. J. Miserendino and R. M. Pick, *Lasers in Dentistry* (Quintessence, Chicago, 1995).
[15] D. Fried, W. Seka, R. Glena, and J. Featherstone, ''Thermal response of hard dental tissues to 9- through 11-μm CO_2 laser irradiation,'' Opt. Eng. **35**(7), 1976–1984 (1996).
[16] I. R. Spears, ''A three-dimensional finite element model of prismatic enamel: a reappraisal of the data on the Young's modulus,'' J. Dent. Res. **76**(10), 1690–1697 (1997).
[17] I. L. Shannon and J. B. Keuper, ''Microhardness of human dentin: baseline values and effects of fluorides.''
[18] V. Lertchirakarn, J. E. A. Palamara, and H. H. Messer, ''Anisotropy of tensile strength of dentin,'' J. Dent. Res. **80**(2), 453–456 (2001).
[19] S. Lees, ''Ultrasonic measurement of thin layers,'' IEEE Trans. Sonics Ultrason. **SU-18**(2), 81–86 (1971).
[20] V. Lertchirakarn, J. E. A. Palamara, and H. H. Messer, ''Anisotropy of tensile strength of root dentin,'' J. Dent. Res. **80**(2), 453–456 (2000).
[21] F. C. Yanikoglu, F. Ozturk, O. Hayran, M. Analoui, and G. K. Stookey, ''Detection of natural white spot caries lesions by an ultrasonic system,'' Caries Res. **34**, 225–232 (2000).
[22] J. D. Aussel and J. P. Monchalin, ''Precision laser-ultrasonic velocity measurement and elastic constant determination,'' Ultrasonics **27**, 165–177 (1989).
[23] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements* (University Science, Sausalito, CA, 1997).
[24] W. S. Brown, W. A. Dewey, and H. R. Jacobs, ''Thermal properties of teeth,'' J. Dent. Res. **49**(4), 752–755 (1970).